

Software for the Collaborative Editing of the Greek New Testament



Troy Andrew Griffiths

Department of Theology and Religion

University of Birmingham

This thesis is submitted for the degree of

Doctor of Philosophy

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This project was responsible for developing the Virtual Manuscript Room Collaborative Research Environment (VMR CRE), which offers a facility for the critical editing workflow from raw data collection, through processing, to publication, within an open and online collaborative framework for the Institut für Neutestamentliche Textforschung (INTF) and their global partners while editing the Editio Critica Maior (ECM)-- the paramount critical edition of the Greek New Testament which analyses over 5600 Greek witnesses and includes a comprehensive apparatus of chosen manuscripts, weighted by quotations and early translations. Additionally, this project produced the first digital edition of the ECM. This case study, transitioning the workflow at the INTF to an online collaborative research environment, seeks to convey successful methods and lessons learned through describing a professional software engineer's foray into the world of academic digital humanities. It compares development roles and practices in the software industry with the academic environment and offers insights to how this software engineer found a software team therein, suggests how a fledgling online community can successfully achieve critical mass, provides an outsider's perspective on what a digital critical scholarly edition might be, and hopes to offer useful software, datasets, and a thriving online community for manuscript researchers.

Dedication

Naturally this work is dedicated to “our great God and Savior, Jesus Christ,” the Father Who “made Him Who knew no sin to be sin on our behalf, so that we might become the righteousness of God in Him,” and the “Spirit Who gives us life!”

([Titus 2:13](#); [2 Corinthians 5:21](#); [Romans 8:2](#))

Acknowledgements

There are way too many people to list here who have helped make this community collaboration a success, though most of them are mentioned in [section 1.6](#) and [2.1.3](#).

1. Introduction	11
1.1. Roadmap through this work	13
1.2. What is critical editing?	16
1.3. Limitations of the printed critical edition	21
1.4. A New Testament textual criticism primer	25
1.5. A developing landscape from 2010	28
1.5.1. Collaborating ECM institutes	30
1.5.1.1. INTF	30
1.5.1.2. ITSEE	31
1.5.1.3. ISBTF	33
1.5.1.4. MOTB Scholars Initiative	34
1.5.1.5. CNTTS	35
1.5.1.6. CSNTM	36
1.5.2. Relevant humanities efforts	36
1.5.2.1. Project Bamboo	37
1.5.2.2. Oxford VRE	38
1.5.2.3. Interedition	38
1.5.2.4. The Huygens Instituut and eLaborate	39
1.5.2.5. The Maryland Institute for Technology in the Humanities and the Text-Image Linking Environment (TILE)	40
1.5.2.6. T-PEN	41
1.5.2.7. Scripto	42
1.5.2.8. TextLab	43
1.5.2.9. Transcribe Bentham	43
1.5.2.10. Perseus	44
1.5.2.11. TextGrid	45
1.5.2.12. Classical Text Editor	46
1.5.2.13. Tyndale House Research Library	47
1.5.2.14. CrossWire Bible Society	47
1.5.2.15. DARIAH, CLARIN, DiXiT	48
1.5.3. The Virtual Manuscript Room (VMR) 1.0 at the INTF	49
1.6. Finding a software team in the humanities	55
2. Community	67
2.1. History	67
2.1.1. Crisis turned catalyst	71
2.1.2. Garnering associates and interest	74
2.1.3. The trek to Münster	76

2.2. A project with a goal	80
2.3. A project with life	81
2.4. Easy level of entry	83
2.5. A project with a soul	85
2.6. Attribution / credit collaborators	85
2.7. Natural contribution	86
3. Design	88
3.1. VMR 1.0: a critical evaluation	89
3.1.1. Frontend / user interface	91
3.1.2. Datastore	93
3.1.2.1. Organize by type, not content	94
3.1.2.2. HSSListe table	100
3.1.2.2.1. Entity Attribute Value (EAV) design	107
3.1.2.3. Besitzer table	112
3.1.2.4. The Majority Tables	114
3.1.3. System structure	119
3.1.4. What happened?	120
3.2. A way forward	123
3.2.1. A normalized data model	124
3.2.2. The benefits of a component architecture in humanities	125
3.3. 4-Tier architecture	126
3.4. Technology selection	131
3.4.1. UTF-8 character encoding	131
3.4.2. TEI	135
3.4.2.1. WYSIWYM	137
3.4.3. Git revision control system	139
3.4.4. MySQL relational database management system	142
3.4.5. Java	143
3.4.6. Web services API	144
3.4.6.1. AJAX and CORS	145
3.4.6.2. REST	146
3.4.6.3. The VMR CRE web services	147
3.5. User interface	148
3.5.1. Use of HTML widgets	149
3.5.2. Inter-Gadget communication	153
3.6. Tagging features	155
3.6.1. RDF	158

3.7. Rights and roles	159
3.8. Scope	160
3.8.1. Use cases	161
3.9. Referenceability	166
3.9.1. Direct object views by persistent URL	166
3.9.2. Queries by persistent URL	168
3.9.3. Data by persistent URL	171
3.9.4. Alignment tables from persistent URLs	172
3.9.5. Variant graphs from persistent URLs	173
3.9.6. Workspaces which honor persistent URLs	174
3.10. Multilingual	176
4. Stages of editing a critical edition	177
4.0.1. My Work Assignments gadget	178
4.0.2. Project Status gadget	179
4.1. Cataloguing Manuscripts	180
4.1.1. The Kurzgefaßte Liste	180
4.1.2. The concept of a manuscript identifier	181
4.1.2.1. The Pinakes Project	182
4.1.2.2. The ParaTexBib Project	183
4.1.3. The Catalogue gadget	187
4.2. Choosing witnesses	198
4.3. Imaging resources	203
4.3.1. Image storage	205
4.3.2. The Page and Image Management gadget	208
4.3.2.1. Pairing Images with Manuscript Pages	209
4.3.2.2. Viewing permissions and image accreditation	211
4.3.2.3. Generating appropriate viewing formats	213
4.3.3. Image viewing	214
4.3.4. Page and Image data model	217
4.4. Indexing folios of manuscripts	218
4.4.1. The Indexing gadget	219
4.4.2. Indexing data model	223
4.4.3. History of indexing tools in the VMR CRE	224
4.5. Transcribing	228
4.5.1. Transcription gadget	230
4.5.2. Transcription reconciliation	231
4.5.2.1. Reconciliation gadget	231

4.5.3. Transcription viewer gadget	234
4.6. Regularizing and Collating	236
4.6.1. Regularization	236
4.6.2. Collation and its history	238
4.6.2.1. Collate	239
4.6.2.2. CollateX and Interedition.	240
4.6.2.3. CollateX for transcription reconciliation	242
4.6.3. Collation and Regularization gadgets	244
4.7. Editing an Apparatus	246
4.7.1. Setting Variant Units	249
4.7.2. Re-ordering variant unit readings	250
4.7.3. Collation and Apparatus Editor gadget	250
4.7.4. Versional and citation evidence	252
4.8. Re-evaluating the editorial reading	252
4.8.1. Components for choosing the initial reading	255
4.8.1.1. Pre-Genealogical Coherence statistical analysis	255
4.8.1.2. Local stemmata	257
5. Open programmatic access	258
5.1. Details of the API	260
5.1.1. Authorization / Rights and Roles	261
5.1.1.1. auth/session	261
5.1.1.2. auth/hasrole	262
5.1.2. Project and group management	262
5.1.2.1. projectmanagement/project	262
5.1.2.2. projectmanagement/usergroup	263
5.1.2.3. projectmanagement/task	264
5.1.2.4. projectmanagement/project/data	265
5.1.3. Metadata	265
5.1.3.1. metadata/v11n (versification)	265
5.1.3.2. metadata/institute	270
5.1.3.3. metadata/manuscript	271
5.1.3.4. metadata/shelfinstance	272
5.1.3.5. metadata/liste (manuscript catalog)	272
5.1.4. Object features	275
5.1.4.1. feature/definition	276
5.1.4.2. feature	277
5.1.5. bibliography	278

5.1.6. image	278
5.1.7. biblicalcontent	279
5.1.8. transcript	279
5.1.9. regularization	281
5.1.10. documentgroup	282
5.1.11. collate	283
5.1.12. Communication forum	286
5.1.12.1. forum/category	286
5.1.12.2. forum/message	287
5.1.13. Test passages	287
5.1.13.1. variant/segment	287
5.1.13.2. variant/reading	288
5.1.13.3. variant/apparatus	288
5.1.14. statistics	289
5.2. Adoption	290
6. Beyond print	296
6.1. The digital edition of the Editio Critica Maior	297
6.1.1. Restoring context	300
6.1.2. A transparent conclusion	302
6.1.3. Community: enhancement and perpetual revision	306
6.1.4. A dynamic research tool	309
7. Summary of developments	318
7.1. Contribution to the field	318
7.2. Future considerations	322

1. Introduction

This project's goal, to construct a turnkey software platform which enables humanities scholars to collaborate together in research and to publish their independent findings, as well as a full digital scholarly edition, has been an "elusive holy grail"—so much so that an entire specialty which focuses largely on this task has been named and accepted in humanities academia, Digital Humanities, despite the lack of such a distinction in other disciplines: the non-existence of Digital Physics, Digital Chemistry, Digital Maths, Digital Law, or Digital Medicine. Why is applying technology to the humanities such a daunting task? Could it be that the study and focus of humanities often places in the limelight the *transmission medium* which today's digital technology actually seeks to deprecate: the written book?

Certainly challenging peculiarities specific to software development in the humanities, and in academia in general, bring their difficulties. Many projects get caught up attempting to solve these peculiar and difficult issues at the expense of missing solid solutions for the mundane but necessary tasks common to most disciplines. This research seeks to marry an established and mature online collaboration platform with solutions custom-developed to meet the challenges in humanities which are elusive to good technology solutions and practices, and to summarily present a turnkey software

framework, along with a case study of its design principles and development history, which navigates the peculiar problems in the world of digital humanities.

This software development research project and case study seeks to convey successful methods and lessons learned while it chronicles a professional software engineer's foray into the world of academic digital humanities. What might be found useful within, beyond the software produced, are insights to how this software engineer found a software development team in the academic environment, suggestions to help a fledgling online community successfully achieve critical mass, an outsider's perspective on what a digital critical scholarly edition might be and become, and some useful datasets curated by a new and thriving online community of New Testament textual scholars.

Specifically, this project is a case study to transition the workflow at the Institut für Neutestamentliche Textforschung (INTF) and their global partners to an open and online collaborative research environment for producing and publishing the *Editio Critica Maior* (ECM)-- the paramount critical edition of the Greek New Testament which analyses over 5600 Greek witnesses and includes a comprehensive apparatus of chosen manuscripts, weighted by quotations and early translations.¹ This transition will establish as a fully accessible and transparent online research dataset the resources of the witnesses to the New Testament gathered in the work to produce volumes of the

¹ The publisher's project page includes the most current data on the volumes completed, to date; Deutsche Bibelgesellschaft, "The Editio Critica Maior," Academic-Bible, 2018, <http://www.academic-bible.com/en/home/current-projects/editio-critica-maior-ecm/>.

ECM. After the workflow is encapsulated and is migrated to an online community, the ECM itself is transitioned into an online digital scholarly edition.

1.1. Roadmap through this work

Before turning to the primary task, this introduction will first give a brief summary of New Testament editions and electronic editing and will then place the project in the context of other related humanities efforts, including a review of the landscape at the time the research began, the institutions focused on critically editing the New Testament and the tools that they use, concluding with an overview of current trends and ongoing development in the wider discipline of digital humanities.

Centrally, this research project builds and documents the research and development of the Virtual Manuscript Room Collaborative Research Environment (VMR CRE) software suite, designed to accommodate electronic editing at the premier institute for New Testament textual research, the Institute für Neutestamentliche Textforschung (INTF). It presents an investigation into and the documentation of the workflow for research, editing, and publishing used in their work to produce, to date, the most exhaustive printed critical Greek New Testament, the *Editio Critica Maior* (ECM). This thesis will describe the author's work alongside the researchers at the Institute to capture and document their workflow as they direct collaborating partners and manage their own teams working to produce fascicles of this edition, specifically with their ongoing work in parallel to finish Acts, to edit the Greek apparatus for the Gospel of

Mark, and to begin transcribing the Gospel of Matthew, their collaboration with the Institute for Textual Scholarship and Electronic Editing (ITSEE) at the University of Birmingham on the Gospel of John, and the Institut für Septuaginta und biblische Textforschung (ISBTF) in Wuppertal with their work on the Apocalypse. Along with documenting the ECM workflow, this study discusses the design, development, and deployment, in collaboration with the INTF, of a bespoke software solution which a) facilitates their workflow in a collaborative online research environment; b) moves this printed work into the realm of digital availability, providing readers with a recontextualized apparatus, traceability from apparatus to transcription to manuscript image, and immediate access to the real-time current state of the dataset used for the edition; c) accommodates the field and community of New Testament manuscript research, establishing a meeting point and research facility to use and contribute to, as they use, a collaborative community-maintained research dataset. A further objective will be to design this software solution to make a general contribution to the academic community-- to facilitate other projects wishing to engage in a similar workflow for digital scholarly editing. Finally, it will discuss what a critical edition might become in future generations.

Central to this project is the element of building an online collaborative community. Lessons learned from past projects along with the tools designed to achieve and retain a thriving collaborative community of researchers will be covered in [Chapter 2](#). A conceptual design for a comprehensive software solution will be presented in [Chapter](#)

[3](#), including guiding principles and clear objectives. A critique of the digital methods and software tools in place at the INTF before this current research project will additionally show the value of this effort. [Chapter 4](#) dissects the critical editing workflow at the INTF into the following 8 discrete stages, with subsections describing each stage along with the challenges therein, and finally describes in detail how the components of the developed software system facilitate the specific editing stage:

- cataloging witnesses, [4.1](#),
- determining which witnesses will be included in the edition, [4.2](#),
- imaging each artifact, [4.3](#),
- indexing what text content resides on each portion of the artifact, [4.4](#),
- transcribing those witnesses, [4.5](#),
- collating and regularizing the text of the transcriptions, [4.6](#),
- editing an apparatus, [4.7](#), and
- re-evaluating the base editorial reading, [4.8](#),

These software components have been developed as needed, or chosen from existing tools and integrated together into a seamless online environment for collaborative research and editing of the Greek New Testament, which now facilitates the Institute's ongoing work, together with their partners and the open community of New Testament researchers. To assure open access for other parties interested in the data produced during the editing process, a web services application programming interface (API)

enables dissemination (see API [Adoption, section 5.2](#)) of these resources ([Chapter 5](#)). Finally, [Chapter 6](#) will consider and address the limitations of the traditional printed edition. In an attempt to push this field beyond the printed scholarly edition, this project will also develop and present an online critical edition facility which will include research tools for the reader to re-contextualize each variation, show full traceability from variation through regularization to transcription to manuscript image, allow the choices made for the edition to be adjusted, including manuscript sets and regularization rules, and always present to the reader the real-time, current state of evidence as they utilize this new online publication of the edition in their study. [Chapter 7](#) will present the adoption of this research, discuss lessons learned, give concluding thoughts and implications, and suggest a way forward.

1.2. What is critical editing?

Critical editing in the Arts and Humanities has a long and established tradition. From editions of the works of Dante, to the United States' Declaration of Independence, textual criticism reveals the history of a work's textual transmission through the study of variance in manuscript copies over time and sometimes even the creative process of an author through the study of surviving autographic revisions. This discipline is not new to the modern era, but continues a practice found in antiquity. Scribes of old often used multiple exemplars while copying and correcting their work, sometime annotating when they met a variation between sources. Many ancient Greek scribes and scholars practiced the preservation and annotation of variants, as Leonard Muellner describes:

“Checking the text against a standard version and preserving rather than suppressing variants was the regular practice, for instance, of the most famous Homeric researcher in the Hellenistic (2nd Century BCE) period in Alexandria, Aristarchus of Samothrace.”² Diglots³ such as Codex Bezae, and Origen’s Hexapla, which aligned 6 and later 8 editions of the Old Testament into columns-- the 5th column being Origen’s Greek version of the LXX-- displayed a keen interest in and laid the foundation for establishing critical editions based on multiple manuscripts.

Efforts toward printed editions of the Greek New Testament began in the early sixteenth century, first in Spain with Cardinal Ximenes’ commissioned Complutensian Polyglot, which was completed on January 10, 1514, but publication was delayed until after the creation of the accompanying Old Testament (July 15, 1517) and approval by Rome on March 28, 1522.⁴ Erasmus started his labors in Basel soon after Cardinal Ximenes and was the first of the two to publish: his edition appeared first in February, 1516, with four more editions following. These two initial publications started a landslide of other works-- often comparing published editions with each other, some citing manuscript evidence. The Stephanus edition of 1550 was the first to include a printed apparatus. Robert Stephanus’ subsequent edition of 1551 first included the New Testament versification as we generally know it today. The seventeenth century

² Leonard Muellner, “Annotations and the Ancient Greek Hero: Past, Present, and Future,” *Comunicar* XXII, no. 44 (January 1, 2015): 46–53, <https://doi.org/10.3916/C44-2015-05>.

³ A diglot is a presentation of a work showing two text in different languages of the same work usually rendered side by side.

⁴ Kurt Aland and Barbara Aland, *The Text of the New Testament: An Introduction to the Critical Editions and to the Theory and Practice of Modern Textual Criticisms* (Leiden: Eerdmans and Brill jointly, 1987), 3f.

begins with the last of five editions from Theodore Beza and editions from the press of the brothers Elzevir-- these texts generally following that of Stephanus. By this time, the idea of the "Received Text" of the Greek New Testament was forming, as first suggested in an Elzevir introduction, "Thou hast the Text now received by all, in which we give nothing altered or corrupted."⁵ This tradition became the standard Greek edition for a century. A renewed interest in variation and manuscript research stirred in the eighteenth and nineteenth centuries, seeing critical editions from Tischendorf (1849), Tregelles (1857), Westcott-Hort (1881), and Weymouth (1892).⁶

While the transmission history of the Greek New Testament-- along with that of critical editions which seek to capture elements of such-- is vast, the path of history which is pertinent for this research project turns to the hand editions of the Greek New Testament of Eberhard Nestle, which was first published in 1898...

"In 1898 Eberhard Nestle published the first edition of his *Novum Testamentum Graece*. Based on a simple yet ingenious idea it disseminated the insights of the textual criticism of that time through a hand edition designed for university and school studies and for church purposes. Nestle took the three leading scholarly editions of the Greek New Testament at that time by Tischendorf, Westcott/Hort and Weymouth as a basis. (After 1901 he replaced the latter with Bernhard Weiß's 1894/1900 edition.) Where their textual decisions differed from each

⁵ Willett L. Adye, *The History of the Printed Greek Text of the New Testament, with the Materials Available for Its Revision Considered: Being a Lecture Delivered at the Hartley Institution, Southampton, Jan. 30th, 1865, with a Supplement* (London: Rivingtons, 1865), <https://archive.org/stream/historyofprinted00adyerich#page/n11/mode/2up>;

Christopher De Hamel, *The Book: A History of the Bible* (London: Phaidon Press, 2005), 220ff.

⁶ A digitized full publication can be found at: <https://archive.org/details/resultantgreekte00weym>.

other Nestle chose for his own text the variant which was preferred by two of the editions included, while the variant of the third was put into the apparatus.”⁷

But it was not until Eberhard’s son Erwin published the 13th edition in 1927 that the apparatus was expanded to cite individual manuscripts, patristic citations, and early translations as evidence for a reading. The 25th edition (1963) gained Kurt Aland as chief editor, taking over the work and fundamentally changing the focus of research away from using other editions as sources, toward examining the ancient source materials, including early papyri and the consideration of all known witnesses up to the twentieth century. While this aim was comprehensive, only a basic survey of the entire body of material was possible. Aland first began his work at the University of Münster in 1958 where he founded the Institut für Neutestamentliche Textforschung (INTF), publishing a catalog of known Greek New Testament witnesses as the *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments*.⁸ Computers were just beginning to be utilized for humanities research⁹ and Aland introduced them into his work toward establishing a critical edition of the Greek New Testament. He began digitally recording data which would initiate a series of volumes to publish the results

⁷ Deutsche Bibelgesellschaft, “History of the Nestle-Aland Edition,” Nestle-Aland Novum Testamentum Graece, 2017, <http://www.nestle-aland.com/en/history>.

⁸ Kurt Aland et al., *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments* (Berlin: De Gruyter, 1963; second edition 1994).

The current, most up-to-date version of the *Liste* is now online using the software developed from this present research project, at: INTF, “*Liste*,” New Testament Virtual Manuscript Room (NTVMR), 2018, <http://ntvmr.uni-muenster.de/liste>.

⁹ Roberto Busa S.J., *Sancti Thomae Aquinatis Hymnorum Ritualium Vara Specimina Concordantiarum: A First Example of Word Index Automatically Compiled and Printed by IBM Punched Card Machines*, vol. N. 7, Serie II (Milan: Fratelli Bocca, 1951).

of his hand collations¹⁰ of a large number of manuscript witnesses at chosen *Teststellen* (= test points)-- volumes of the *Text und Textwert*.¹¹ While this printed work composes a valuable compilation of a cross-section of the tradition, when the research for this thesis was begun in 2010, there existed no publicly available electronic dataset of the manuscript evidence for the Greek New Testament. While catalogs, images, collations and other materials have been accumulated by Aland and his institute, much of it remained in analog form stored in filing cabinets and on microfilm. Facilitating the transition of these valuable resources to an open and transparent, community-maintained digital repository is a primary objective for the VMR CRE software and this research project.

The Nestle-Aland *Novum Testamentum Graece* is at its 28th edition at the time of this writing, representing a summary of the research carried on by the institute begun by Aland, in Münster, Germany; their principal focus is the expanded work of the *Editio Critica Maior* (ECM)¹² of the Greek New Testament. This is the primary Institute and project with which we will collaborate to achieve the goal of this thesis-- a publicly-available and community-maintained dataset of manuscript evidence for the Greek New Testament, achieved through the development of end-to-end software to support the collaborative editing and research of the same, adding knowledge to how

¹⁰ A collation is a list of differences between a manuscript and a printed base text (David C. Parker, *An Introduction to the New Testament Manuscripts and Their Texts* (Cambridge, UK: Cambridge University Press, 2008), 95.

¹¹ Kurt Aland et al., Series: *Text und Textwert der griechischen Handschriften des Neuen Testaments* (Berlin: De Gruyter, 1987-), <https://www.degruyter.com/view/serial/16713>.

¹² Institut für Neutestamentliche Textforschung, "Academy Project 'Novum Testamentum Graecum - Editio Critica Maior,'" INTF, http://egora.uni-muenster.de/intf/aecm/aecm_en.shtml.

the presentation of and research toward a critical edition is re-invented in the global digital era.

1.3. Limitations of the printed critical edition

This past decade has seen an explosion in interest from the humanities for a desire to move the art of critical text research and editing to the digital realm. This is true in the domain of New Testament text criticism as well.¹³ Vast amounts of scattered data have made their way online, including searchable manuscript catalogs, holding institution image repositories, electronic texts of Greek and Latin works from antiquity, and more. This brings to the New Testament scholar an exponential increase in material to be studied, the accessibility of new mediums for research, the expectation for transparency in research sources, and immediate access to the latest updates of the materials-- all of which poses fundamental problems for the traditional print format. As Ulrich Schmid notes, “the analogue printed scholarly New Testament edition faces one main challenge, namely the unwieldiness of the sheer amount of data: Greek manuscripts, versional and Patristic data. The efforts to cope with this challenge result in an apparatus with the following major drawbacks: de-contextualization of the evidence, lack of traceability, non-updateability.”¹⁴ The printed critical edition has evolved in creative ways to overcome the limitations of its physical medium of transmission. The first edition of

¹³ Hugh A.G. Houghton and Catherine J. Smith, “Digital Editing and the Greek New Testament,” in *Ancient Worlds in Digital Culture* (Leiden: Brill, 2016), 110–27, doi:10.1163/9789004325234_007.

¹⁴ Ulrich Schmid, “Transmitting The New Testament Online,” in *Text Comparison and Digital Creativity*, ed. Ernst Thoutenhoofd, Wido Th. Peursen, and Adriaan Weel (Leiden: Brill, 2010), 189–206, doi:10.1163/ej.9789004188655.i-328.73.

५३५

Stephanus, First ed. From Acts

became a challenge for the printed medium, as we see from a page of the *Editio Critica*

*Maior*¹⁵ below.

[illegible]

ECM, James 2:3b

This page from the ECM presents only 17 words from the Greek text along the top, with the remainder of the page recording the support for variant readings. The printed

¹⁵ Barbara Aland et. al., eds., *Novum Testamentum Graecum. Editio Critica Maior / Die Katholischen Briefe: Teil 1 und Teil 2, Gesamtwerk* (Stuttgart: Deutsche Bibelgesellschaft, 2013).

page constrains the edition to a small fragment of the work's text per page and the overwhelming majority of the apparatus body is confined to symbols, abbreviations, and sigla, which either assumes an expert reader in both the field and this edition, or requires the reader to constantly flip between the pages of the edition body, a table of symbols, and a list of sigla. When new evidence is found or errors in the edition are discovered, many years often pass before these are published. The fascicle containing the second half of this verse from James, shown above, was first published in 1997¹⁶ with the first update not coming until a second edition released in 2013.¹⁷ All of the vast data used to compile this apparatus-- an individual witness' transcription in context or an image of the folio containing this verse, regularization rules applied, or the legibility of the extant text itself, etc.-- is not easily accessible if the reader wishes to research more deeply into an apparatus entry or confirm the statements made in the edition. The print medium prohibits accessibility to the entire dataset which might otherwise be made available via a digital delivery mechanism, allowing not only the current generation of researchers the ability to confirm the results in the edition, but also enabling the next generation of scholars to build on the massive amount of research done for the edition. Digital tools offer not only freedom from the many constraints of the printed page for publishing the end product of an edition project, but also they have the potential to better facilitate collaboration between disparate teams working on such large edition projects and to offer opportunities for an extended

¹⁶ Barbara Aland et al., eds., *Novum Testamentum Graecum. Editio Critica Maior IV / Die Katholischen Briefe: Teil 1, 1. Lieferung, Der Jakobusbrief* (Stuttgart: Deutsche Bibelgesellschaft, 1997).

¹⁷

Barbara Aland et al., eds., *Novum Testamentum Graecum. Editio Critica Maior / Die Katholischen Briefe: Teil 1 und Teil 2, Gesamtwerk* (Stuttgart: Deutsche Bibelgesellschaft, 2013).

community of researchers to participate in the process.

1.4. A New Testament textual criticism primer

The New Testament is the most copied, most printed, and most studied body of text in human history.¹⁸ Even so, when the present research project began in 2005, there was little in the way of public digital data of the manuscript evidence to the New Testament. Some data could be harvested from web pages, and projects like the Early Greek Bible Manuscripts Project under the direction of Peter Head at Tyndale House in Cambridge were beginning to collect resources.¹⁹ This body of manuscript evidence is abundant, with approximately 2.1 million pages catalogued of continuous text and lectionary Greek New Testament witnesses²⁰ and these materials are often difficult to access, even for the manuscript scholar willing to travel. These artifacts are housed at universities and museums scattered across the globe and the vast majority of these resources await digitization. Fortunately, the INTF has a comprehensive microfiche collection of images for almost every known New Testament manuscript.

¹⁸ Standard introductions to New Testament textual criticism include: Kurt Aland and Barbara Aland, *The Text of the New Testament: An Introduction to the Critical Editions and to the Theory and Practice of Modern Textual Criticisms* (Leiden: Eerdmans and Brill jointly, 1987).

David C. Parker, *An Introduction to the New Testament Manuscripts and Their Texts*, 1 edition (Cambridge, UK: Cambridge University Press, 2008).

Bart D. Ehrman and Michael W. Holmes, *The Text of the New Testament in Contemporary Research: Essays on the Status Quaestionis*. Second Edition (Leiden: Brill, 2012).

¹⁹ See this snapshot of the project page from as early as 2002: Peter M. Head, "Early Greek Bible Manuscripts Project," Tyndale House, 2002, <http://web.archive.org/web/20020116073727/http://www.tyndale.cam.ac.uk/Tyndale/staff/Head/EGBMP.htm>.

²⁰ The Institut für neutestamentliche Textforschung, "Statistics - Pages," NT.VMR, 2017 <http://ntvmr.uni-muenster.de>.

The discipline of New Testament textual criticism focuses on researching the 5,691 known extant Greek manuscripts which give witness to the New Testament, as catalogued in November 2016 in the online *Kurzgefaßte Liste*.²¹ This significant volume of witnesses from diverse countries, copied by hand over sixteen centuries, for different intended audiences, will have some variation among the copying tradition. A primary objective for the New Testament textual critic is to establish the initial text²² of the New Testament by examining all the variation in the textual tradition. The production of this initial text is often accompanied by a set of notes showing where any important witnesses deviate from the chosen wording of the initial text; this is called a *variant apparatus* (or *critical apparatus*). The final product is called a *critical edition of the Greek New Testament*. The wording of the biblical text used in the edition is considered an *eclectic text*, i.e., a form established not from a single document witness, but from evaluating many document witnesses and constituting a text based on a selection from the body of evidence.²³

Diverse specialized fields of research have developed hand in hand with the practice of textual criticism and the production of scholarly editions. Palaeography is the study of handwriting and helps decipher and is one means to help date manuscript witnesses.

²¹ The Institut für neutestamentliche Textforschung, “Manuscript Workspace,” NT.VMR, 2017, <http://ntvmr.uni-muenster.de/community/vmr/api/metadata/liste/search/?docID=10001-49999>

²² A comprehensive definition of the term *initial text*, per the Institut für Neutestamentliche Textforschung can be found in the section on Perspective in the Introduction to the *Editio Critica Maior*, available online at: The Institut für neutestamentliche Textforschung, “Perspective,” The *Editio Critica Maior* (ECM), 2017 <http://www.uni-muenster.de/INTF/ECM.html>

²³ For a sample of different editing methodologies, see Chapters 23, 26, & 27, from Bart D. Ehrman and Michael W. Holmes, *The Text of the New Testament in Contemporary Research: Essays on the Status Quaestionis. Second Edition* (Leiden: Brill, 2012).

The fields of codicology and papyrology research the materials and processes involved in copying a text along its history of transmission.²⁴ These disciplines have helped establish the order and placement for leaves within a manuscript, assisting in reconstituting a manuscript from dispersed fragments, and have contributed to the rediscovery of lost biblical witnesses from *palimpsest* parchment-- pages scraped and reused for other writings. Even the biological science of Phylogeny is tapped for techniques which assist in grouping manuscript witnesses into genealogies.²⁵ The New Testament manuscript tradition provides a challenging genealogical puzzle due in part to the scribal practice of referencing multiple sources while producing a new copy; this practice was sometimes used to ensure high accuracy of the resultant copy or simply due to the compound nature of the New Testament, with the Gospels, Pauline letters, Acts and Catholic Epistles, and the Apocalypse often circulating in discrete early compilations; thus, a commissioned complete New Testament might need as many as 4 early exemplars to supply a complete source text. This cross-pollination of variants in a single manuscript originating from different genealogical lines of manuscripts is referred to as genealogical contamination.²⁶ Variation in a particular manuscript cannot always be attributed to a singular ancestral line, and thus other complex genealogical research methods have been developed for studying the specific New Testament

²⁴ Parker includes a very concise, yet informative four page history of Palaeography, Papyrology, and some Codicology in his preface to his article, David C. Parker, "Was Matthew Written before 50 CE?: The Magdalen Papyrus of Matthew," *The Expository Times* 107, no. 2 (November 1995): 40–43, <https://doi.org/10.1177/001452469510700203>. For a more detailed outline, see Parker, *An Introduction*.

²⁵ Adrian C. Barbrook et al., "The Phylogeny of The Canterbury Tales," *Nature* 394, no. 6696 (August 27, 1998): 839–839, <https://doi.org/10.1038/29667>.

²⁶ Matthew Spencer, Klaus Wachtel, and Christopher J. Howe, "The Greek Vorlage of the Syra Harclensis," *TC: A Journal of Biblical Textual Criticism* 7 (2002), <http://rosetta.reltech.org/TC/v07/SWH2002/index.html>.

manuscript tradition.²⁷ New Testament textual criticism is rich in innovative history and has contributed back to the advancement of many of these fields. For a full introduction to New Testament textual criticism, see David Parker's *An Introduction to the New Testament Manuscripts and their Texts*.²⁸

1.5. A developing landscape from 2010

In New Testament textual criticism and scholarly editing-- more specifically, those working on the *Editio Critica Maior* (ECM), there were three institutes hosting projects and two additional strategic partners who contribute to this work, when this project officially started as research, in 2010. Since that time, one new institute has begun preliminary work on an ECM project. These six institutions are dispersed across the world with their primary focus some aspect of text-critical research on the Greek New Testament. These institutions will be the target user community for our tools and are:

- Institut für Neutestamentliche Textforschung (INTF), Münster, DE,

²⁷ See, Gerd Mink's Coherence-Based Genealogical Method.

Gerd Mink, "Eine Umfassende Genealogie Der Neutestamentlichen Überlieferung," *New Testament Studies* 39, no. 4 (October 1993): 481–99, <https://doi.org/10.1017/S0028688500011929>.

———, "Problems of a Highly Contaminated Tradition: The New Testament. Stemmata of Variants as Source of a Genealogy for Witnesses," in *Studies in Stemmata II*, ed. Pieter Th van Reenen, A. A. den Hollander, and Margot van Mulken (Amsterdam: John Benjamins Publishing, 2004), 5–85.

———, "Contamination, Coherence, and Coincidence," in *The Textual History of the Greek New Testament: Changing Views in Contemporary Research*, ed. Klaus Wachtel and Michael W. Holmes (Atlanta, GA: Society of Biblical Literature, 2011), 141–205.

Tommy Wasserman and Peter J. Gurry, *A New Approach to Textual Criticism: An Introduction to the Coherence-Based Genealogical Method* (Atlanta: Society of Biblical Literature, 2017), <http://www.jstor.org/stable/j.ctt1whm8zr>.

²⁸ Parker, *An Introduction*.

- Institute for Textual Scholarship and Electronic Editing (ITSEE), Birmingham, UK,
- Institut für Septuaginta und Biblische Textforschung (ISBTF), Wuppertal, DE,
- Museum of the Bible Scholars Initiative (MOTB), OK, US,
- Center for New Testament Textual Studies (CNTTS), LA, US, and
- Center for the Study of New Testament Manuscripts (CSNTM), TX, US.

It is important to note that these institutes already collaborate together both by traditional means and also electronically in an ad hoc manner. ITSEE is the home of the work for the International Greek New Testament Project (IGNTP) , collaborating with the INTF and the ISBTF on the *Editio Critica Maior*. The MOTB Scholars Initiative sponsors and promotes research teams investigating Bible artifacts, including New Testament Manuscripts, with multiple teams contributing transcription work toward ongoing and potential future ECM work. In the CNTTS work to produce manuscript collations²⁹ their team supports the IGNTP with manuscript indexing data. CSNTM is primarily concerned with amassing a library of digital images which may be shared with the other projects. These organizations make up the target collaborative community for our project. We will focus on discovering and meeting the challenges these organizations have as they share in work together. A brief background for each of these institutions follows.

²⁹ Published as a module in the Accordance Bible Software product.

1.5.1. Collaborating ECM institutes

1.5.1.1. INTF

The Institut für Neutestamentliche Textforschung is located in Münster, Germany and has the commission “to research the textual history of the New Testament and to reconstruct its Greek initial text on the basis of the entire manuscript tradition, the early translations and patristic citations.”³⁰ True to this mandate, they have since 1959³¹ primarily compiled collations as they examine manuscripts. The INTF has collated the vast majority of extant Greek NT manuscripts for selected readings throughout the New Testament and this data can be found in the *Text und Textwert* volumes.³² Only within the past decade has the preferred mode for recording the content of a manuscript witness shifted from paper collation to full electronic transcription.³³ Kurt Aland, the founder of the Institute, produced a comprehensive list of all known continuous text Greek New Testament witnesses and published this as *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments*.³⁴ The INTF is recognized by all leading institutions as the authority for the continued curation of this authoritative catalogue (hereafter referred to as the *Liste*), and in 2010 the INTF began to make this

³⁰ Institut für Neutestamentliche Textforschung, “INTF,” Universität Münster, 2017, http://egora.uni-muenster.de/intf/index_en.shtml.

³¹ Institut für Neutestamentliche Textforschung, “History,” INTF, 2017, http://egora.uni-muenster.de/intf/institut/geschichte_en.shtml

³² Kurt Aland et al., *Text und Textwert der griechischen Handschriften des Neuen Testaments* (Berlin: De Gruyter, 1987).

³³ Ulrich B. Schmid, W. J. Elliott, and David C. Parker, *The Gospel According to St. John: Majuscles* (Leiden: Brill, 2007), 3.

³⁴ Kurt Aland et al., *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments* (Berlin: De Gruyter, 1963, second edition 1994), <http://public.eblib.com/choice/publicfullrecord.aspx?p=3042078>.

list available online along with image and transcription resources.³⁵ For more information on cataloguing manuscripts and the *Liste*, [see Section 4.1](#). Current projects underway at the INTF include the *Editio Critica Maior*, Digital Nestle Aland, NT Transcripts, and the Virtual Manuscript Room.³⁶ The *Editio Critica Maior* (ECM),³⁷ referred to above, is a long term effort to re-evaluate the full New Testament manuscript tradition. The INTF has entered into formal partnership with IGNTF and ISBTF on this work. The research work on the ECM provides a catalyst for our project to work together assisting the institutes involved.

1.5.1.2. ITSEE

David Parker and Peter Robinson joined forces in 2005 to form the Institute for Textual Scholarship and Electronic Editing at the University of Birmingham in the United Kingdom.³⁸ While much exciting work in Digital Humanities is ongoing at ITSEE, the project most pertinent to our topic is the ECM efforts for the Gospel of John. This work takes place under the auspices of the International Greek New Testament Project (IGNTF).³⁹ It has been located at ITSEE thanks to externally-funded projects led by David Parker, Executive Editor of the IGNTF for the Gospel according to John since 1988. ITSEE staff were largely responsible for producing transcriptions from images,

³⁵ Institut für Neutestamentliche Textforschung, “*Liste*,” NT.VMR, 2017, <http://ntvmr.uni-muenster.de/liste/>.

³⁶ Institut für Neutestamentliche Textforschung, “Project of the INTF,” INTF, 2017 http://egora.uni-muenster.de/intf/projekte/projekte_en.shtml

³⁷ For interesting insights into the methods and motives for the ECM see, David C. Parker and Klaus Wachtel, “The Joint IGNTF/INTF *Editio Critica Maior* of the Gospel of John: Its Goals and Their Significance for New Testament Scholarship” (Annual Meeting of SNTS (Studiorum Novi Testamenti Societas), Halle, August 2, 2005), <http://epapers.bham.ac.uk/754/>.

³⁸ Institute for Textual Scholarship and Electronic Editing, “History,” ITSEE, 2010, <https://web.archive.org/web/20100727143709/http://www.itsee.bham.ac.uk/history.htm>

³⁹ International Greek New Testament Project, “IGNTF,” IGNTF, <http://www.igntp.org>

and linked alignment between the transcription and the images for the Codex Sinaiticus Project⁴⁰ bringing together the British Library,⁴¹ the National Library of Russia,⁴² St. Catherine's Monastery,⁴³ and Leipzig University Library.⁴⁴ This monumental work has provided ITSEE staff with extensive experience launching a sizable digital edition. The present research project collaborated with Hugh Houghton, drawing from his experience with the Codex Sinaiticus Project, to assist him in crafting a suitable TEI schema⁴⁵ for capturing the transcription of a New Testament witness, starting by introducing him to the Roma TEI schema generator and later by advising when any questions arose regarding the intended usage of a particular TEI construct.⁴⁶ In 2011, Rachel Kevern and Bruce Morrill served the ECM John project, managing a team of transcribers with an aim to produce digital transcriptions of New Testament manuscripts where there was need for their work on the Gospel of John and also for a planned future edition of Galatians. They have an established process in place for managing transcription work and have produced training materials adequate for bringing up to par the palaeography skills of a would-be volunteer with a reasonable

⁴⁰ Codex Sinaiticus Project, "Codex Sinaiticus, Experience the Oldest Bible," Codex Sinaiticus, 2017, <http://codexsinaiticus.org>

⁴¹ British Library, "The British Library, For Research, Inspiration, and Enjoyment," The British Library, 2017, <http://www.bl.uk>

⁴² National Library of Russia, "The National Library of Russia, The Pride of Russian Culture," The National Library of Russia, 2017, <http://nlr.ru/eng/>.

⁴³ Holy Monastery of St. Catherine at Mount Sinai, "Mount Sinai Monastery," The official site of Sinai Saint Cathrine Monastery, 2017, <http://www.sinaimonastery.com>

⁴⁴ Universitätsbibliothek Leipzig, "Universitätsbibliothek Leipzig," UBL, 2017, <http://www.ub.uni-leipzig.de>

⁴⁵ ITSEE, "TEI-NTMSS," WCE-OTE, 2017, <https://sourceforge.net/p/wfce-ote/code/ci/master/tree/wfce-ote/TEI-NTMSS.xsd>

⁴⁶ H. A. G. Houghton, "The Electronic Scriptorium: Markup for New Testament Manuscripts," in *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, ed. Claire Clivaz, Andrew Gregory, and David Hamidović (Leiden: Brill, 2013), 31–60, <http://www.brill.com/products/book/digital-humanities-biblical-early-jewish-and-early-christian-studies>.

background.⁴⁷ In 2011 they were utilizing a shared spreadsheet for managing transcription assignments, a server folder where they can drop images to be downloaded by volunteers, and email for exchanging transcription work completed.

1.5.1.3. ISBTF

The Institut für Septuaginta und Biblische Textforschung (ISBTF), Wuppertal, Germany was founded in 2009 although this effort has older roots preparing work on the Septuagint from 1995. At that time, a University network was formed which included Prof. Dr. Martin Karrer at the Kirchliche Hochschule Wuppertal/Bethel, Prof. Dr. Wolfgang Kraus, Universität Koblenz and Landau, and later Universität des Saarlandes, Saarbrücken (New Testament). In 1999 this network started the successful German translation of the Septuagint.⁴⁸ In 2006, the international conferences for Septuagint Studies were organized which continue to take place every second year. The congress volumes are published by Mohr Siebeck Tübingen. In 2007, Martin Karrer started a Deutsche Forschungsgemeinschaft (DFG) funded project on Septuagint studies researching the Septuagint quotations in the New Testament which finished in the beginning of 2012, with Siegfried Kreuzer (Old Testament) focusing on the Antiochene text of the Septuagint beginning in 2009 and ongoing up to 2016. In 2009, the Hochschule coordinated these studies as the ISBTF. In 2011, the *Editio Critica*

⁴⁷ Rachel Kevern, "Transcribing Greek Minuscule Manuscripts," Monograph, March 23, 2011, <http://epapers.bham.ac.uk/774/>.

⁴⁸ Wolfgang Kraus and Martin Karrer, eds., *Septuaginta Deutsch: Das griechische Alte Testament in deutscher Übersetzung*, 2nd ed. (Stuttgart: Deutsche Bibelgesellschaft, 2009); Martin Karrer and Wolfgang Kraus, *Septuaginta Deutsch - Erläuterungen und Kommentare: Band 1 und 2*, 1st ed. (Stuttgart: Deutsche Bibelgesellschaft, 2011).; Siegfried Kreuzer, *Handbuch zur Septuaginta: Einleitung in die Septuaginta* (Gütersloh: Gütersloher Verlagshaus, 2016); Eberhard Bons and Jan Joosten, *Handbuch zur Septuaginta: Die Sprache der Septuaginta* (Gütersloh: Gütersloher Verlagshaus, 2016).

Maior Apokalypse Project began, sponsored again by the DFG. From 2011-2015 Ulrich Schmid (Church history) worked with the ISBTF and also at the INTF and has partnered as the primary domain expert to help direct the production of the software for the research project of this paper. When work began on the ECM *Apokalypse* edition at the ISBTF, the VMR CRE research tools under development had already progressed to a point which could facilitate their work, and they are the first project to start at the beginning of their workflow using the VMR CRE framework.

1.5.1.4. MOTB Scholars Initiative

“As the academic research wing of Museum of the Bible, the Scholars Initiative fosters biblical research at colleges, universities, and seminaries across the world, planning and supporting academic projects related to the languages and material culture of the Bible, and capitalizing on artifacts in the Museum Collections.”⁴⁹ Initially started as a means to research the tens of thousands of biblical resources privately amassed by the Green family, owners of the 550+ store Hobby Lobby chain, the MOTB Scholars Initiative has matured into a research network focused on raising up the next generation of scholars on biblical artifacts. As world-class experts are employed to research items in the collections at the museum, students are assigned to each research team, proactively encouraging their academic development. Each year, these students have the opportunity to participate in the Logos summer conference, most recently held at the University of Oxford in England, where they experience lectures and workshops given by leading experts in their field. The Scholars Initiative is important for this research

⁴⁹ Museum of the Bible, “Scholars Initiative,” MOTB, 2018, <https://www.museumofthebible.org/research/scholars-initiative>.

project, as they have one team working toward the publication of Greek Paul manuscripts and a critical edition in collaboration with the IGNTP as part of the *Editio Critica Maior*. They have materially contributed to the ongoing support and development of the VMR CRE and actively use the software for their research.

1.5.1.5. CNTTS

Led by William Warren, the H. Milton Haggard Center for New Testament Textual Studies⁵⁰ is a relatively young institute, founded in 1998 as a research center of the New Orleans Baptist Theological Seminary. They offer specialized training and degree courses for students wishing to pursue the study of the Greek New Testament manuscripts. They seek to develop digital resources and research aids for scholars in the field, and are intentional about cooperating on projects at other institutes. One of this institute's most significant projects, to date, is the CNTTS Critical Apparatus which is available as a module for the major Bible software projects and allows a researcher to search variants from a large number Greek New Testament manuscripts. CNTTS provides a talented pool from which to draw scholars who have the qualifications necessary to produce digital resources from ancient New Testament manuscripts.

⁵⁰ New Orleans Baptist Theological Seminary, "The H. Milton Haggard Center for New Testament Textual Studies," NOBTS, <http://www.nobts.edu/CNTTS>

1.5.1.6. CSNTM

Daniel Wallace directs the Center for the Study of New Testament Manuscripts with the primary purpose to “make digital photographs of extant Greek New Testament manuscripts so that such images can be preserved, duplicated without deterioration, and accessed by scholars doing textual research.”⁵¹ CSNTM has a team of experts negotiating permission to photograph New Testament witnesses at holding institutions worldwide. Regular practice for CSNTM is to offer their professional photography and digitization services to holding institutions at no cost in exchange for access to the manuscripts, concluding with both CSNTM and the holding institution benefitting from a new, quality, high resolution image set at no cost to the latter. Their efforts have amassed high resolution digital photographs preserving, as of this writing in 2016, 844 Greek New Testament witnesses⁵² and this collection is growing steadily. The first collaborative community effort for the VMR CRE was in partnership with the CSNTM to collect indexing data for their image collection. Details of this history and how the tools and relationship evolved can be found in [section 4.4](#) with special reference to the history in [subsection 4.4.3](#)

1.5.2. Relevant humanities efforts

Before beginning this research and development, a survey of existing tools and efforts was undertaken along with an attempt to open channels of communication and establish

⁵¹ Center for the Study of New Testament Manuscripts, “About the CSNTM,” CSNTM, 2017, <http://csntm.org/home/about>

⁵² Center for the Study of New Testament Manuscripts, “Manuscript Search,” CSNTM, 2017, <http://www.csntm.org/Manuscripts.aspx>

partnerships across relevant projects. While not the case for all projects reviewed, it is evident now, concluding this research project, that many of the efforts investigated at the start of this research and described in the following sections have failed to attain their stated goals or to produce any usable software for critical editing and manuscript research. A common theme observed in *post-mortem* assessments by project participants seems to center around lack of clear unified purpose or the identification of a need in the non-digital humanities world which would welcome a solution-- ending almost as an effort in search of a goal, or a goal in search of a user base. This has been a crucial observation for how best to avoid this same fate. Having a pilot project (=the ECM editorial teams and their workflow) to drive goals and development efforts has given a concrete finish line to steer toward-- even while keeping a potentially wider audience in mind during development and design.

1.5.2.1. Project Bamboo

Funded by the Andrew W. Mellon Foundation starting in 2008 and continuing through 2012, Project Bamboo “aimed to enhance arts and humanities research through the development of infrastructure and support for shared technology services. Its planning phase brought together scholars, librarians, and IT staff from a wide range of institutions, in order to gain insight into the scholarly practices Bamboo would support, and to build a community of future developers and users for Bamboo’s technical deliverables. From its inception, Bamboo struggled to define itself clearly and in a way that resonated with scholars, librarians, and IT staff alike. The early emphasis on a service-oriented architecture approach to supporting humanities research failed to

connect with scholars, and the scope of Bamboo's ambitions expanded to include scholarly networking, sharing ideas and solutions, and demonstrating how digital tools and methodologies can be applied to research questions... the lack of a shared vision that could supersede the individual interests of partner institutions resulted in a scope around which it was difficult to articulate a clear narrative. When Project Bamboo ended in 2012, it had failed to realize its most ambitious goals."⁵³

1.5.2.2. Oxford VRE

The Building a Virtual Research Environment for the Humanities (BVREH) efforts at Oxford had at least two rounds of funding for projects very similar to this present research project; however, at the time of this writing, no usable software has yet been produced. The first effort was only intended to be a demonstration project, the Virtual Workspace for the Study of Ancient Documents.⁵⁴ The second, A VRE for the Study of Documents and Manuscripts states that the project intends to build OpenSocial research gadgets, similar to this project. Collaboration could have occurred, but efforts to contact the participants in this effort and later, those participating at Oxford in the Bamboo effort listed previously, yielded no interest from their teams.

1.5.2.3. Interedition

The Interedition effort was particularly instrumental toward developing ideas and relationships that have lasted throughout the life of this project. Funded by the

⁵³ Quinn Dombrowski, "What Ever Happened to Project Bamboo?," *Literary and Linguistic Computing* 29, no. 3 (September 1, 2014): 326–39, <https://doi.org/10.1093/lc/fqu026>.

⁵⁴ Ruth Kirkham, "EPSRC/AHRC Funding for e-Science Demonstrator Project," Building a Virtual Research Environment for the Humanities, 2017, http://bvreh.humanities.ox.ac.uk/news/e-Science_Demonstrator.html.

European Union between 2008 and 2012, COST Action IS0704: An Interoperable Supranational Infrastructure for Digital Editions (Interedition)⁵⁵ set out to bring together a number of humanities departments across European universities. Many participating in this effort had previously produced digital tools for their own individual purposes, with large overlap, both with each other and with New Testament studies. The Interedition project intended to increase collaboration between institutions and to unify these toolsets. The first and primary component which came from this effort was a baseless collation software engine, CollateX,⁵⁶ which has seen adoption by many projects within Interedition and without. For more on CollateX and the concept of baseless collation, [see Section 4.6.2.2](#). Interest to participate in Interedition extended beyond the EU. Some relevant projects at the time, from participating Interedition members, included:

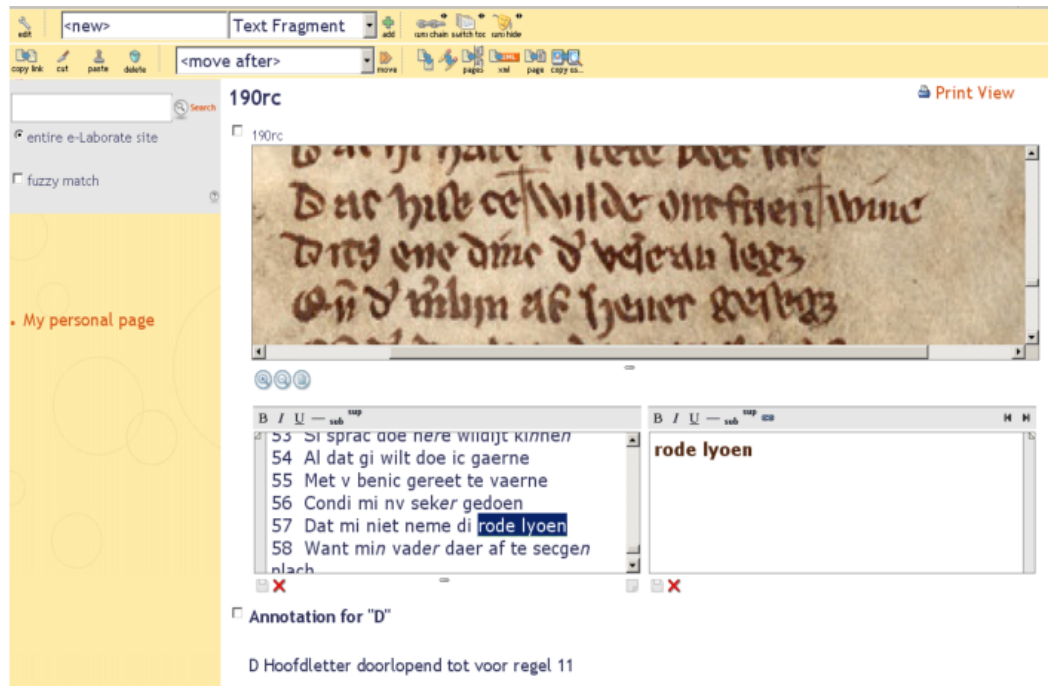
1.5.2.4. The Huygens Instituut and eLaborate

The Huygens Instituut, KNAW is a research institute for text editions and textual scholarship of the Royal Netherlands Academy of Arts and Sciences. One product they have produced is eLaborate, a web-based tool for the preparation of digital text-editions. It is written in Java and includes a transcription editor and a digital edition publisher.⁵⁷

⁵⁵ Interedition, “ISCH COST Action IS0704,” European Cooperation in Science and Technology, 2017, http://www.cost.eu/COST_Actions/isch/IS0704.

⁵⁶ Ronald Haentjens Dekker et al., “Computer-Supported Collation of Modern Manuscripts: CollateX and the Beckett Digital Manuscript Project,” *Literary and Linguistic Computing* 30, no. 3 (September 1, 2015): 452–70, <https://doi.org/10.1093/lc/fqu007>; Interedition Development Group, “CollateX – Software for Collating Textual Sources,” CollateX, 2017, <http://collatex.sourceforge.net>

⁵⁷ Huygens Instituut KNAW, “e-Laborate: Virtual Workspace for Social Sciences and Humanities,” e-Laborate, 2017, <http://www.e-laborate.nl>



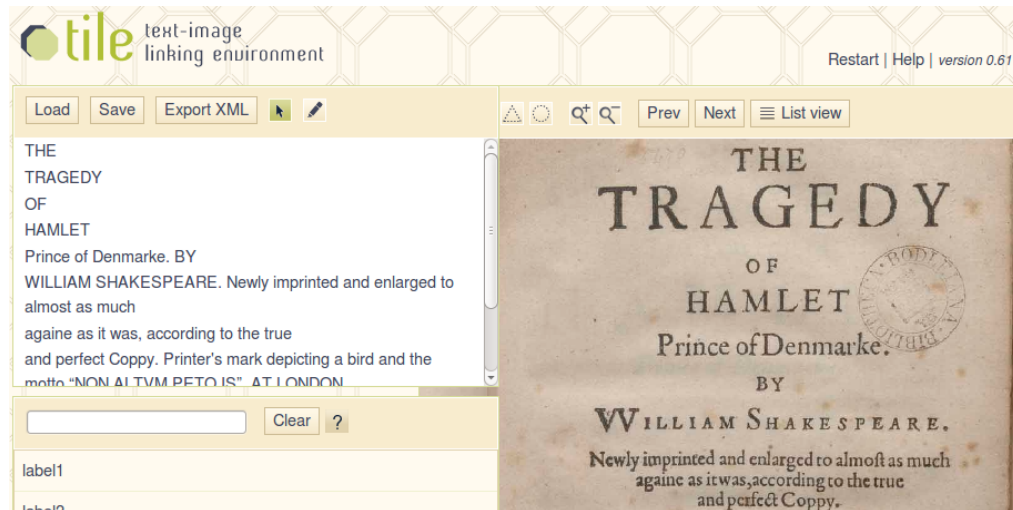
e-Laborate

Ronald Dekker and Joris van Zundert of the Huygens Instituut, who helped develop e-Laborate have lent their experiences through the Interedition collaboration to contribute to the success of the CollateX (for CollateX, see [4.6.2.2](#)) engine and for the overall organization of the Interedition project. Initial collaboration with Joris van Zundert and Tara Andrews produced the initial incarnations of what has become a variant graph collation view and regularization rule input tool for the VMR CRE (see sections [4.6. Regularizing and Collating](#)).

1.5.2.5. The Maryland Institute for Technology in the Humanities and the Text-Image Linking Environment (TILE)

From the Maryland Institute for Technology in the Humanities, TILE aimed to be a collaborative image markup tool for both manual and semi-automated linking between

encoded text and images, and was developed mostly in ECMAScript (a.k.a., JavaScript):⁵⁸



TILE

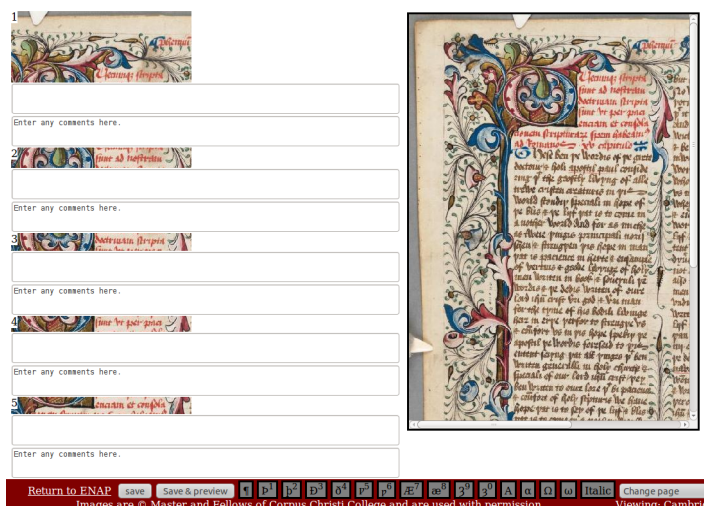
TILE was written under the direction of Doug Reside, and while the TILE project is now defunct, Doug reviewed and contributed advice and lessons learned from his efforts on TILE toward the input for design on the VMR CRE. Doug is now applying his experience as the Digital Curator for the Performing Arts at New York Public Library.

1.5.2.6. T-PEN

Transcription for Paleographical and Editorial Notation is a tool under development by Saint Louis University to enable community-based, line-by-line transcription and annotation. Work was originally funded by the NEH and the Andrew W. Mellon Foundation. The project launched in 2012. Images of manuscripts must be uploaded to

⁵⁸ Doug Reside, "TILE: Text Image and Linking Environment," TILE, 2011, <https://web.archive.org/web/20110822061224/http://mith.info/tile>.

the website. The tools will then attempt to automatically identify lines of text within a scanned page.⁵⁹ An edit box immediately under each line in the manuscript image provides a means for a user to transcribe. Any desired markup must be entered by hand in the transcription line editor. The project claims to have 1500 unique users working on 2000 projects. A call for new suggestions and desired features was issued in 2015.⁶⁰ No code has been contributed between that time and the time of this writing.⁶¹



T-PEN

1.5.2.7. Scripto

From The Center for History and New Media at George Mason University, Scripto is a lightweight, open source tool that will allow users to contribute transcriptions to online documentary projects. Scripto is attempting to abstract the transcription tools away

⁵⁹ James Ginther, "T-PEN's Project Management," Center for Digital Theology at Saint Louis University, 2011, <http://digital-editor.blogspot.com/2011/03/t-pens-project-management.html>

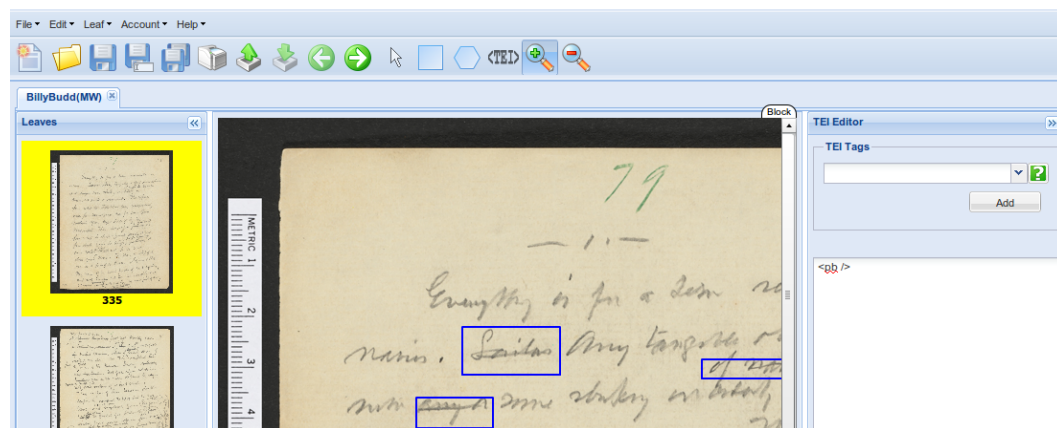
⁶⁰ Patrick Cuba, "T-PEN Development Advance Post," The Walter J. Ong, S.J., Center for Digital Humanities, 2015, <https://blog.ongcdh.org/development/t-pen-development-advance-post/>.

⁶¹ Saint Louis University Center for Digital Humanities, "Github Contributors to TPEN3," TPEN3, 2017, <https://github.com/CenterForDigitalHumanities/TPEN3/graphs/contributors>

from the repository functionality in an effort to allow institutions to interface with their existing datastores.⁶²

1.5.2.8. TextLab

From Computing Services at Hofstra University, TextLab is a tool for assisting transcription of manuscript images using the TEI specification. It is developed with ECMAScript and Flash.⁶³



TextLab

1.5.2.9. Transcribe Bentham

A participatory project based at University College London, Transcribe Bentham is aimed to engage the public in the online transcription of original and unstudied

⁶² Roy Rosenzweig Center for History and New Media, "Scripto, A Community Transcription Tool," Scripto, 2017, <http://scripto.org>

⁶³ Performant Software Solutions, "About TextLab," TextLab, 2017, <http://www.textlab.org/about/>.

manuscript papers written by Jeremy Bentham. The Transcribe Bentham project is implemented by extending the MediaWiki collaboration framework⁶⁴



Transcribe Bentham

1.5.2.10. Perseus

The Perseus Project is one of the more successful projects to produce open data for ancient literature. With a mandate to “make the full record for humanity as intellectually accessible as possible to every human being, providing information adapted to as many linguistic and cultural backgrounds as possible,” they have indeed digitized and made accessible vast amounts of data since their inception in 1987. At

⁶⁴ Louise Seaward, “Transcribe Bentham, A Participatory Initiative,” UCL Transcribe Bentham, 2017, <http://blogs.ucl.ac.uk/transcribe-bentham/>.

the time of this writing, Perseus at Tufts shows approximately 168 million words of material available. A key benefit to this project from Perseus has been the motivation to other institutions to provide texts which work with Perseus' software in a tightly defined TEI markup called EpiDoc. The use of TEI for this current research project is heavily influenced by EpiDoc.

1.5.2.11. TextGrid

TextGrid began in 2006 as an effort to produce a virtual research environment to offer “scholars in the humanities sustainable editing, storing and publishing of their data in a thoroughly tested and safe environment.”⁶⁵ Built on the Eclipse Platform, the TextGrid Laboratory is a software solution which consists of a set of OSGi⁶⁶ components providing DH functionality running within the Eclipse Framework, together offering a research environment which is downloaded and runs locally on a researcher's workstation-- connecting to an online server for user authentication. TextGrid also has amassed from partner contribution a collection of works from over 600 German authors, offered as the Digital Library of TextGrid and more generally, with ongoing additional contributions as the TextGrid Repository.⁶⁷ While this technology attempted to solve many of the intricacies pertinent to our efforts, key contributors to TextGrid have expressed their desire for a web-based solution and have partnered with us through the Workspace for Collaborative Editing to build the Online Transcription

⁶⁵ TextGrid, “TextGrid: A Project and its History,” TextGrid Project, 2017, https://textgrid.de/en_US/projekt.

⁶⁶ OSGi Alliance, “The Dynamic Module System for Java, OSGi, <https://www.osgi.org>.

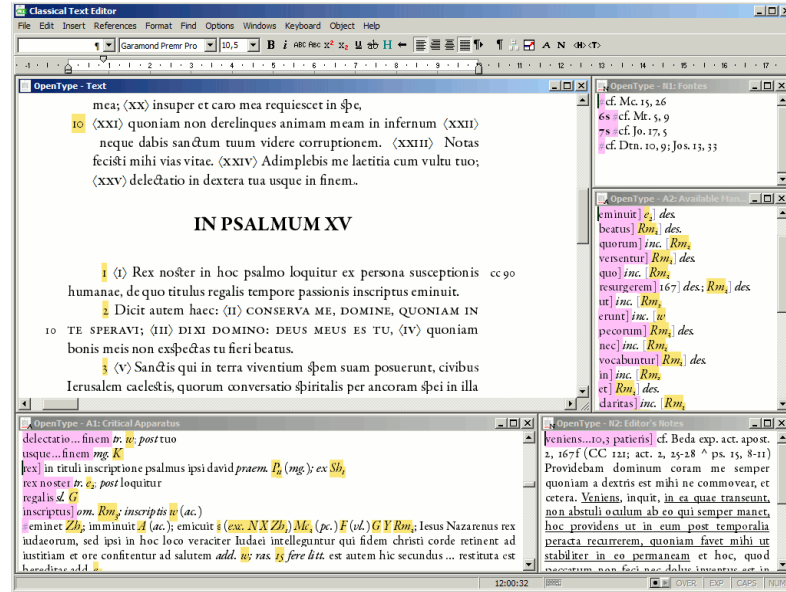
⁶⁷ TextGrid, “TextGrid Repository,” TextGrid Repository, 2017, https://textgridrep.org/en_US/repository.html.

Editor component of the VMR CRE. During 2012-2015, the German Federal Ministry of Education and Research (BMBF) funded the final segment of development on TextGrid and concluded with efforts to migrate TextGrid into the DARIAH-DE consortium.

1.5.2.12. Classical Text Editor

The Classical Text Editor filled a need for many years, and still does today, serving as a specialized word processor for authoring digital editions. From the website, The Classical Text Editor is a native Windows program which claims to be “the word-processor for critical editions, commentaries and parallel texts,” supporting “any number of notes and apparatus - bidirectional text - OpenType - sigla.” Developed by Stefan Hagel in 1997, The Classical Text Editor is an initiative of the Austrian Academy of Sciences and the Corpus Scriptorum Ecclesiasticorum Latinorum (CSEL). Focusing on producing a print edition, “The Classical Text Editor was designed to enable scholars working on a critical edition or on a text with commentary or translation to prepare a camera-ready copy,” or a simple electronic publication. It allows apparatus editing without worrying about page boundaries and also includes some research capabilities.⁶⁸

⁶⁸ Stefan Hagel, “Classical Text Editor: The word-processor for critical editions, commentaries and parallel texts,” Classical Text Editor, 2017, <http://cte.oeaw.ac.at>.



Classical Text Editor

1.5.2.13. Tyndale House Research Library

Tyndale House is a residence library with affiliation to the University of Cambridge network of research libraries. Tyndale hosts and offers housing and research desks for University PhD researchers, scholars on sabbatical, and a variety of other biblical researchers. Some staff hold teaching duties within the Faculty of Divinity. Staff at Tyndale House working on manuscript research have provided invaluable support and feedback for this research project.

1.5.2.14. CrossWire Bible Society

CrossWire⁶⁹ has a history of over 25 years building open source biblical research software. They have produced international standards for the technical exchange of data between Bible societies worldwide under the direction of the United Bible

⁶⁹ CrossWire Bible Society, “The CrossWire Bible Society: Bringing the Gospel to a New Generation,” The CrossWire Bible Society, 2018, <http://crosswire.org>.

Societies,⁷⁰ the Society of Biblical Literature,⁷¹ and the American Bible Society.⁷²

CrossWire has hosted community collaboration projects in the past with good success; however, the tools built for these collaboration projects have not been designed with reuse in mind. Specific tailoring to each project's unique needs has left the resultant tools difficult to reuse.⁷³ Intending to change this deficit, CrossWire is developing a community collaboration and component framework, along with specialized project components, with a view to host a number of future collaborative biblical research projects. For more on CrossWire, see [section 2.1](#).

1.5.2.15. DARIAH, CLARIN, DiXiT

These three large, ongoing collaboration efforts in Europe offer opportunity for the digital researcher to find collaborating partners and discover tools, services, and ideas offered by a large participating body. Digital Research Infrastructure for the Arts and Humanities (DARIAH)⁷⁴ is a support network for digital tools, data, and knowledge transfer. The European Research Infrastructure for Language Resources and Technology (CLARIN), “makes digital language resources available to scholars, researchers, students and citizen-scientists from all disciplines, especially in the humanities and social sciences, through single sign-on access.” Useful is CLARIN’s

⁷⁰ United Bible Societies, “About the United Bible Societies,” UBS, 2017, <https://www.unitedbiblesocieties.org/about-us/>.

⁷¹ Society of Biblical Literature, “Society of Biblical Literature,” SBL, 2017, <http://sbl-site.org>.

⁷² American Bible Society, “The American Bible Society: Giving People Access to the Life-changing Message of God’s Word,” ABS, 2017, <http://www.americanbible.org>.

⁷³ CrossWire Bible Society, “Project KJV2003,” KJV2003, 2017, <http://crosswire.org/sword/kjv2003>.

⁷⁴ DARIAH-EU, “Digital Research Infrastructure for the Arts and Humanities,” DARIAH, 2017, <http://dariah.eu>.

service to curate data from a project. The focus of CLARIN is linguistic data.⁷⁵ Digital Scholarly Editions Initial Training Network (DiXiT) “is an international network of high-profile institutions from the public and the private sector that are actively involved in the creation and publication of digital scholarly editions.” DiXiT offers a great opportunity to hear what other researchers are attempting in the field of digital scholarly editions, and to build connections for sharing in work together. Unfortunately DiXiT was only funded to run through August of 2017.⁷⁶

1.5.3. The Virtual Manuscript Room (VMR) 1.0 at the INTF

Finally, important is the work which had already begun in 2007 at the Institut für Neutestamentliche Textforschung, funded by the Deutsche Forschungsgemeinschaft, to make available online the Greek New Testament manuscript resources used by the INTF to edit the *Editio Critica Maior*. From the initial grant application,⁷⁷ the long term objectives included digitizing for online use the nearly complete microfiche image collection of known New Testament manuscripts housed at the Institute along with images from other online sources, collected bibliographic information on each

⁷⁵ CLARIN ERIC, “CLARIN - European Research Infrastructure for Language Resources and Technology,” CLARIN, 2017, <https://www.clarin.eu>.

⁷⁶ DiXiT, “DiXiT - Digital Scholarly Editions Initial Training Network,” DiXiT, 2017, <http://dixit.uni-koeln.de>.

⁷⁷ “Zusammenfassung: Die Verbesserung des Zugangs zu den Primärquellen der neutestamentlichen Überlieferung ist ein dringendes Desiderat der neutestamentlichen Wissenschaft. In einem virtuellen Handschriftenlesesaal des Instituts für neutestamentliche Textforschung (INTF) sollen eigene digitalisierte Mikrofilmbestände und von anderen Institutionen im Internet publizierte Fotos und Materialien zusammengeführt und mittels Indexierung und Transkription durchsuchbar gemacht werden. Eine ausführliche digitale Handschriftenliste soll die am INTF über viele Jahre erstellten ‘Dossiers’ zu den neutestamentlichen Handschriften der wissenschaftlichen Öffentlichkeit zur Verfügung stellen und weitere Informationen aufnehmen, die auch von nicht dem INTF angehörenden Wissenschaftlern eingebracht werden können.” (provided by Klaus Wachtel, July 2017).

manuscript, and transcriptions of these witnesses. The site would also host the online successor to Aland's printed *Kurzgefaßte Liste*-- the authoritative catalog of Greek New Testament manuscripts, currently maintained at the INTF. The site first went live, November, 2010 and at its peak in late 2011 had the following functionality:



VMR 1.0, Home Page

The initial entry page, shown above, announced news and provided a menu to access the major components of the system, along with buttons to reference other electronic facilities at the Institute: The NTT (New Testament Transcripts), G-Queries (Genealogical Queries), SMR (Coptic New Testament Catalog). As part of the VMR 1.0, the “VMR” button took the reader directly to either of two reading modes, described later. “VMR-Expert” did the same as “VMR” but offered more images restricted for viewing “for research only.” The “Map” functionality was never completed but intended to show the reader where in the world manuscripts currently

reside. The “List” button gave access to the online *Kurzgefaßte Liste*, which provided a facility to retrieve a list of catalogued manuscripts filtered by one of a select criteria, pictured below.

Universität

Münster

Institut

für neuere und mittelalterliche

Textforschung

Handschriftenliste

Select a manuscript by

Object-ID

go

Select a manuscript by

GA-number

go

Select a manuscript by

former GA

go

Select a manuscript by

Country

go

Select a manuscript by

Place

go

Select a manuscript by

Institution

go

Select all manuscripts

with images

go






743 manuscript entities with images in NT.VMR or external image links

412 manuscript entities with images in NT.VMR

ObjID	InstID	GA	Century	Leaves	Address	Images VMR external		
10001	0	P1	III	2 Frg (1 Dbl.)	443	✓		details
10006	0	P6	IV	15 Frg (4 Blatt)	74	✓		details
10009	0	P9	III	1 Frg	405	✓	✓	details

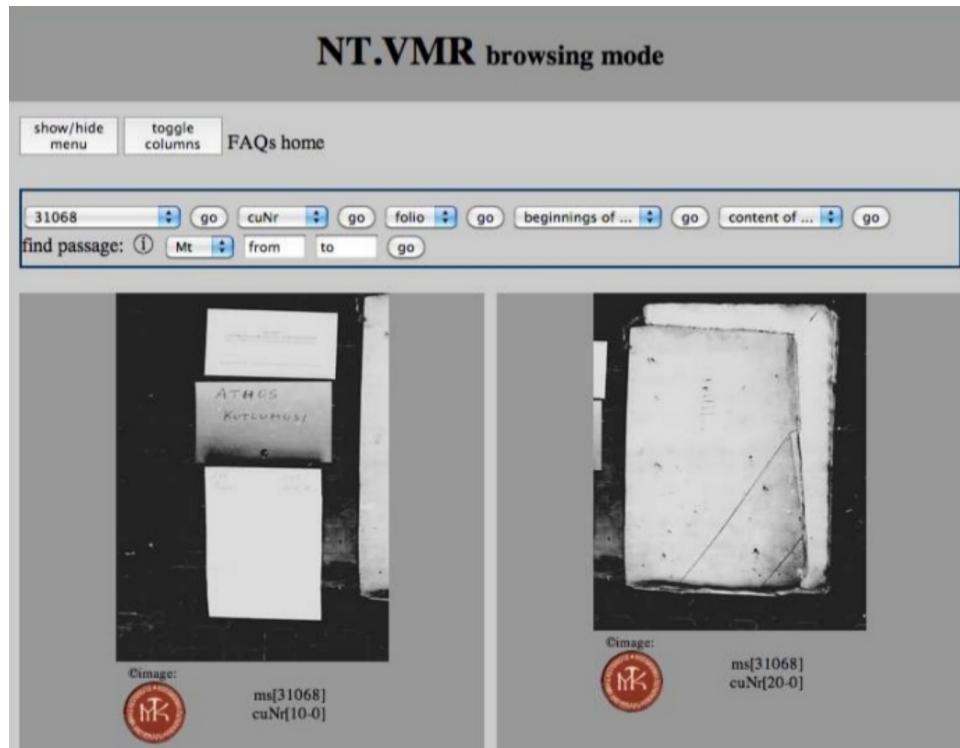
VMR 1.0, Search Page

The details for an individual manuscript included all the main categories of information from the *Kurzgefaßte Liste*.

image  trans  index  FAQs HOME  <input type="button" value="go"/>  <input type="button" value="go"/>	
GA number	P1
Object ID	10001
Instance ID	0
Content	e – Mt 1,1-9.12; 1,14-20.23
Language	g
Century	III
Material	Papyrus
Leaves	2 Frg (1 Dbl.)
Columns	1
Lines	27r, 25v [reconstructed 37-38]
Height	14,7; 2,5 cm [reconstructed 25 cm]
Width	15; 3 cm [reconstructed 13 cm]
Address	<div>443</div> University of Pennsylvania University Museum of Archeology and Anthropology, the Egypt Section, Philadelphia/Pa. (USA)
Shelf number	E 2746
LDAB	2940

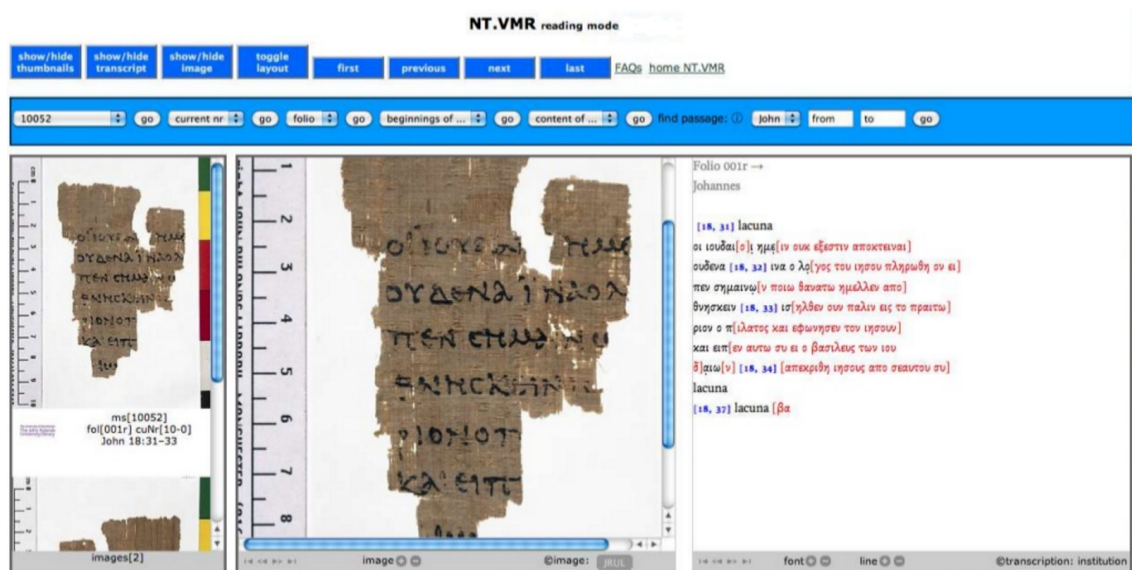
VMR 1.0, Manuscript Details

From here, a reader could proceed to view the images of the manuscript in either of two view options. First, “Browsing Mode” would display at a reduced size every image for a manuscript on a single page which could be scrolled through to select an individual image for full display.



VMR 1.0, Browsing Mode

The second mode for viewing a manuscript was called “Reading Mode” and to the left of the image included a transcription of the page, if available.



VMR 1.0, Reading Mode

Finally, a tool was created to facilitate online indexing contributions from privileged users.

NT.Virtual Manuscript Room - Indizierung

Willkommen

Wir begrüßen Sie herzlich auf der Seite zum Indizieren digitalisierter Handschriftenphotos.

Zur Erfassung der Inhalte digitalisierter Photos haben Martin Faßnacht und Ulrich Schmid ein standardisiertes Verfahren entwickelt, das online über eine einfach zu bedienende Benutzeroberfläche ausgeführt werden kann.

Die Indizierung ist eine wesentliche Voraussetzung zur Entwicklung des Virtual Manuscript Rooms für neutestamentliche Handschriften (NT.VMR), der im Rahmen einer Förderung durch die DFG am Institut für neutestamentliche Textforschung (INTF) aufgebaut wird.

Was ist "Indizieren"?

Unter "Indizierung" versteht man die Erfassung des Inhaltes einer Seite einer Handschrift.

Dazu werden die Handschriften zuerst digital gespeichert: entweder durch das Einscannen vorhandener Mikrofilmaufnahmen, Photos oder Faksimiledrucke oder durch das Erstellen neuer digitalisierter Photos von den Originalhandschriften. Für jedes Digitalisat wird in einer MySQL-Datenbank ein Datensatz angelegt und mit dem entsprechenden Photo verknüpft.

Um eine bestimmte Stelle einer neutestamentlichen Handschrift im NT.VMR direkt aufrufen zu können, muß der Inhalt jeder einzelnen Seite erfasst werden. Dabei wird für das erste und das letzte Wort der Bibelvers bestimmt und im Datensatz gespeichert. Falls die Seite keinen durchgehenden Text bietet (z.B. bei Lektionaren) oder Lücken aufweist, kann für jeden Teiltext einer Seite der Inhalt separat erfasst werden.

Indizierung

ms nr ①
30546

current nr ①
60 - 0

repro nr ①
1

image url
<http://www.ezra.org/Manus>

manuscript view ① on/off reset

folio 002v page 0

☐ illumination

☐ decoration

☐ canon table

☐ list of kephalaia

☐ colophon

☐ index of sections

☐ blank page

☐ other

blook 01

eBook 01

bChap 01

eChap 01

bVers 06

eVers 14

additional view ①

☐ scan status info

☐ microfilm title

☐ note

☐ binding

☐ scale

☐ black scan

☐ other

Optimiert für Firefox & Safari!

Eingabemaske go

Stand der Indizierung go

Mithilfe go

Kontakt go

VMR 1.0, Indexing Instructions

VMR 1.0 was a successful step to initially expose online to a public audience some of the resources available at the INTF. [Chapter 3](#) begins with a critical evaluation of this initial implementation.

1.6. Finding a software team in the humanities

This section serves as a very concise software development primer and also to convey what challenges were met when attempting to find a complete software team in the humanities. It is a reflection of the unique difficulties and also offers insights with how these difficulties were overcome for the success of this project.

Successful software companies, which have no primary motivation other than to be successful software companies, have established strongly developed methods of software development with an organizational model generally common to the field⁷⁸. The humanities, and software development in academia in general, almost always has less capacity, little structure, and has unique influences on software development not always conducive to success. For more on this, see the section following the review of the VMR 1.0, [“3.1.4. What happened?”](#) The plan for this project to succeed, from a software development perspective, was initially to use these same commercial roles and practices-- to identify and utilize the standard players listed below. This was not a straightforward hurdle to clear in the humanities. What follows are the actors involved for a typical small to mid-sized software development company, interleaved with how these roles were satisfied for this project. Some creativity had to be used to fill the

⁷⁸ For software company organizational models, see: Ronald E. Giachetti, *Design of Enterprise Systems: Theory, Architecture, and Methods* (Boca Raton: CRC Press, 2016), 6ff; Mandy Chessell and Harald C. Smith, *Patterns of Information Management* (Upper Saddle River, NJ: IBM Press, 2013), 69ff.

roster in the academic setting. By convention, italics in this section denote what came to be vs. the conventional role description.

2 Technical Sales Representatives - A Technical Sales Representative knows both the domain market which their company's software targets, and also enough technical background to have intelligent conversations with the technical team of a potential client. Their job function is to interact with the general market, discover who plays the role of software purchaser within a potential client organization, pique their interest enough to win a conversation with the technical lead who recommends purchases, and finally convince this person that they cannot live without the services of the software company. They move "potential clients" into the category of "clients" and provide feedback to their organization on what the market desires.

With this collaborative, partially crowd-sourced project, sales solicited a slightly different audience in two ways. First, there was a ready consumer base for New Testament research data, but instead of charging money, this project was asking for precious time; the would-be consumer was asked to commit their own research time within the system to help produce the end product, and that is an expensive price. To sell the idea, Klaus Wachtel, Ulrich Schmid, and this researcher posed the goal and opportunity at many conferences and on blogs and mailing lists popular to the audience.⁷⁹ A common 'sales technique' was to advertise to teachers, "Do you teach

⁷⁹ Carl Conrad, "CSNTM MSS Indexing Project," B-Greek, 2009, <http://lists.ibiblio.org/pipermail/b-greek/2009-July/049724.html>; Jonathan Robie, "Low Hanging Fruit in

Greek? Assign your students 10 pages of manuscript indexing work to help them learn!” This was a unique pedagogical tool teachers could try with their students and this project gained users and data. It was always important to offer some immediate benefit to the potential user. Users were attracted to the initial dataset and would come to visit the site, but when they found a unindexed manuscript which caused them to hunt through the pages to find what was important to them, the offer was made to “index this page.” This was a desired benefit to the user-- so they might find that place again next time they returned, and also moved that user from a consumer to a contributor, naturally.

Next, many of the potential users of parts of the system were internal and also had to be convinced. This is common for a large corporation with an internal programming department, though not usually the responsibility for the independent software development company. Internal potential users of the new software at the INTF, the ISBTF, and ITSEE needed to be sold on the idea of changing their familiar work process. A case had to be made that the new software would actually improve their work experience to the degree that it warranted change from the familiar. This responsibility fell largely on these same three individuals. A key here was to observe and listen to the problems the future users of the software system had with their current workflow, to watch and notice the tedious tasks they did by hand and to offer support,

“If you had the ability to click here, and to see a list of this data, would that make it easier for you?” This latter activity overlapped largely with the next role.

2 Business Analysts (at smaller companies a Technical Sales Representative will serve double-duty) - This role is designed to shield the Software Development Team from the Client (and vice versa). Often less technical than a Technical Sales Representative, the Business Analyst has an expert knowledge of the target domain-- if financial accounting is the domain market, then this person will have extensive experience as an accountant and be well versed in the software solutions for accounting. They interact very well with people. Their function is to spend time with a client, observe their daily workflow and see their needs, recommend enhancements in software to improve their lives, and finally to write “use cases”⁸⁰ and get “sign off” from the client for what will be delivered.

Indeed, with this project, our Technical Sales Representatives played double-duty as Business Analysts. Wachtel and Schmid, both leading scholars in New Testament textual criticism and also ultimately users themselves of the software, knew their colleagues’ workflow, needs and desires, and consulted with them regularly about what tools would make their life easier.

⁸⁰ A use case describes a discrete desired outcome from a software system in a specific scenario. For more on use cases see the start of [section 3](#).

1 Software Architect - This role is ultimately responsible for the technology choices and general design for the company's solutions. The Software Architect is often labeled by Upper Management as the Chief Technology Officer (CTO) and regularly will oversee the daily technical operation of the Software Development Team.

1 Software Engineer - The software engineer translates "Use Cases" from the Business Analyst into "Functional Requirements" for the Software Development Team. They understand how to meet each "Use Case" with a solution already developed by the Software Company or how to expand and improve on the usefulness of an existing solution to cover a new "Use Case". The function of this position is to design a technical plan ("Functional Requirements") to deliver on every promised "Use Case".

3 Software Developers - Typically consisting of 2 Sr. and 1 Jr. developer, the Software Development Team implements the functional requirements established by the Software Engineer. A new university computer science graduate will start in a Jr. role and be mentored by the Sr. developers for a number of years until they understand the ins and outs of the successful software development workplace and master their development skills. A typical career path travels into the Software Development Team from university to Jr. Developer to Sr. Developer, on to Software Engineering, and finally to Architecture. These roles often blend between transition, usually dictated by the size of the company.

Having no other funding for additional software specialist positions for this project, it was necessary to build collaborative relationships with other contemporary efforts. The Workspace For Collaborative Editing project involved Catherine J. Smith in Birmingham, and Martin Sievers in Trier to build two primary components incorporated into this software system.⁸¹ The Interedition⁸² effort by the European Union offered an opportunity to forge partnerships with Joris van Zundert and Tara Andrews which provided a regularization user interface,⁸³ with Ronald Dekker and Gregor Middell which provided the CollateX collation engine, and also importantly, a developer support community where ideas could be exchanged; these partnerships formed the software team for this effort. A lone humanities software developer must reach out to find peers. The temptation to build everything from scratch must be avoided. Build one thing well and find others who have done the same. The majority of needs of a humanities project are not unique to humanities-- every project needs a shared calendar, document store, message forum, wiki, a content management system for an easy to update website, blogs, news. Most software systems need a login facility, rights and roles management. None of these are specific to humanities and there are many good choices available to meet these needs. Do not try to build them. Others have already built them well. This project chose to use a popular open source portal, Liferay, as a base to fill these voids. When specialist tools are needed for a humanities project, first try to attach to an effort already building components which could be used. Make them better. Most importantly identify peers in projects who share like

⁸¹ See the Collation and Apparatus Editor in [section 4.7.3](#) and the Transcription Editor in [section 4.5](#).

⁸² See [section 1.5.2.2](#).

⁸³ See [section 4.6.3](#).

goals. When a component truly is absent, talk about it with these peers and work together. It is notable the importance European-wide efforts like Interedition and later, DiXiT,⁸⁴ play in forging these relationships, without which, many of the partnerships for this project would not have been formed.

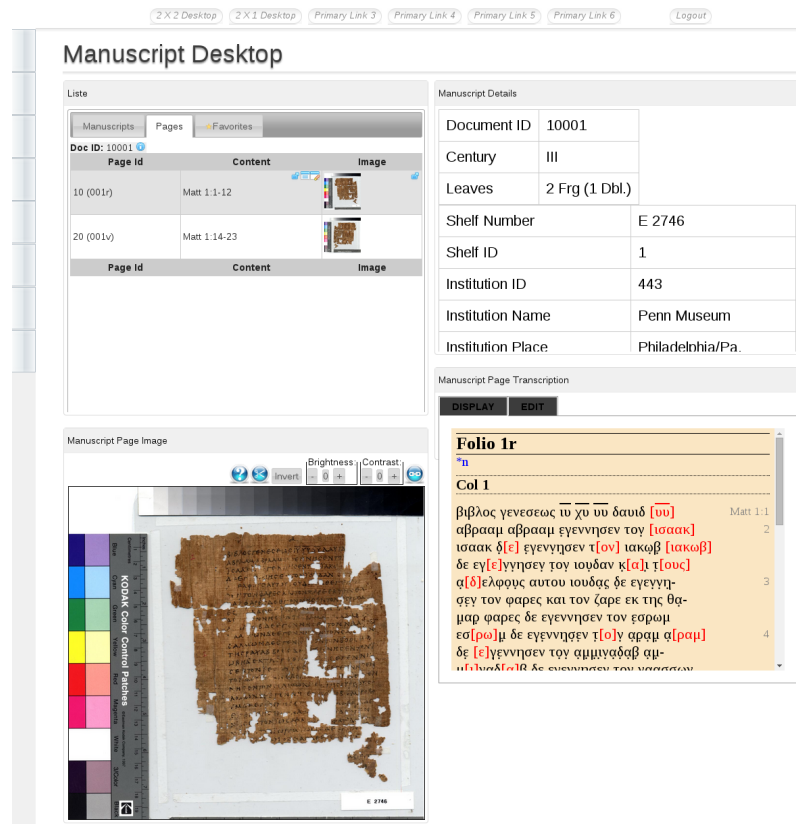
1 Web / UI Designer - The designer is a user interface expert. They have an eye for aesthetics and usability. They are responsible for the “look and feel” of the company’s solutions. They work closely with the Business Analyst to produce “storyboards”⁸⁵ to show how the solution will look and operate, which are given to the client to help them “sign off” on the proposed solution. It is very common for the designer to draw icons, create buttons, and build full page layouts directly used by the Software Developers during implementation.

As mentioned earlier, van Zundert and Andrews provided user interface design and a working prototype for the regularization interface, Smith and Sievers developed the user interfaces for the Collation and Apparatus Editor, and Transcription Editor, respectively. Additionally, Georg Gabel provided evaluation and valuable feedback on the regularization interface. Key was making the software system immediately usable and beneficial to attract early adopters, even if not the most aesthetically pleasing and ergonomic, and to listen to their feedback. Finally, recognizing that UI and graphic

⁸⁴ DiXiT, “DiXiT - Digital Scholarly Editions Initial Training Network,” DiXiT, 2017, <http://dixit.uni-koeln.de>

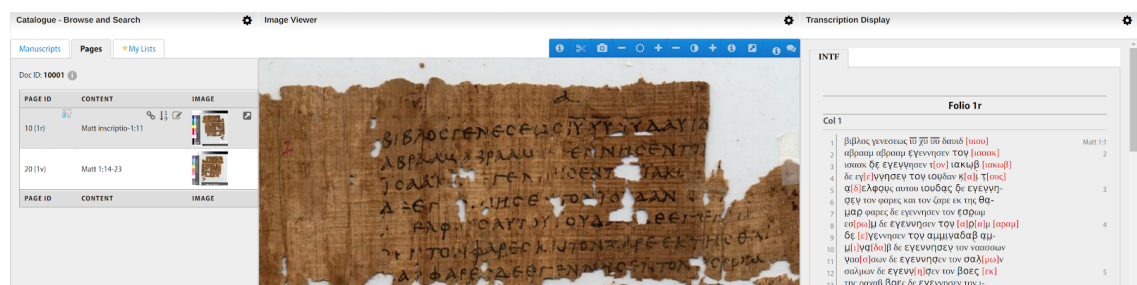
⁸⁵ Scott Stiner, “Software Development Tips: Storyboard For Simplicity,” Forbes, accessed April 9, 2018, <https://www.forbes.com/sites/forbestechcouncil/2017/04/06/software-development-tips-storyboard-for-simplicity/>.

design was a lacking skill of this researcher and that the talent was not present on the local team, solicitation was made early on to the developer community at CrossWire to offer user interface design suggestions, which yielded an influential design from Jonathan Batteas, an established web designer working for a public library in Ohio who graciously donated his time and talents. The result was a visually appealing 2x2 grid of manuscript gadgets with a slide-out sidebar from the left which could discretely house other, less frequently used tools, which could be quickly accessible when the need arose. Many of these same initial UI ideas are incorporated in the software system today.



Original UI idea from Batteas

Additionally, a mechanism was included into each gadget which encouraged end users to supply their own styles, which was used by Peter Gurry, a graphic and web designer turned textual critic, to offer a more aesthetically appealing user interface, shown below.



Gurry's style, near the end of the project

2 Quality Assurance Technicians - The QA team is responsible for a “test plan” to assure that the developed software meets the functional requirements and thus the use cases promised as a solution. They repeatedly execute this test plan and provide feedback to the Software Development Team. They work with Software Engineers and Developers to develop “unit tests” for each component of the system which can be run in the future to guard against regression as the system changes. They typically are good software script writers who may not have a formal Computer Science degree, but have the know-how to get things done.

Unfortunately, without a formal QA team, students and staff at the INTF were exposed to the software at stages before it was ready for general availability. It was key to set expectations for these users that they would indeed be ‘testing experimental software’ while they worked, and to provide a direct channel for their feedback. As a concession, they felt they were given early access to new tools and were contributing their work experience and their preferences to the design. Another benefit to choosing to extend existing tools is the likelihood those tools already had unit tests available. Many of the common Bible functionality for this software was based on libraries with their original genesis in The SWORD Project, from CrossWire, which does have a regression test suite.⁸⁶

⁸⁶ CrossWire Bible Society, “SWORD SVN Test Suite,” SWORD SVN Repository, 2017, <http://crosswire.org/svn/sword/trunk/tests/testsuite/>.

Upper Management and Infrastructure - this includes a Project Manager, Payroll, CEO, CFO, etc. and is common to most successful business models-- not specific to a Software Company.

For more on the software team this project found in academia, see the end of [section 2.1.3](#).

This full software team in the academy is a rare luxury. More typically, a project proposal will include 1 position, sometimes allotted to as low as 25% time working to satisfy the digital needs for a project, and no concession to maintain, much less update the deliverables after the project time (commonly, three years) is complete.

Unfortunately, the salaries attached to these positions-- a fraction of what an experienced engineer would command in the commercial workforce-- reflect the perceived value for an experienced computer scientist. Often an edition will have one textual scholar in the project with a self-taught ability to navigate enough technology working with a university IT department to set up a makeshift workspace to store files, images, and other resources for the project. It is common to find the ‘digital’ aspect of a project weighted heavily by funding bodies, but for the textual scholar, simply an afterthought or worse, a burdensome requirement relegated to the final months of a project plan, and not an enabling, integral component for the success of the research efforts. Two examples come to mind. First, the 100+ year Göttingen

Septuaginta-Unternehmen⁸⁷ effort to produce a critically edited edition of the Greek Old Testament ended their funding in 2016, unable to find a new grant source. They have never published any of their work digitally online and only when they feared losing their funding, hired a Spanish software company to bring a digital presence to their project. The software never materialized and this monumental project is now stalled. The second example has to do with efforts to establish a digital humanities center at the University of Münster, where the INTF is located. Meetings were held with humanities projects at the University who might benefit from a center there, and a common question among the more senior scholars was simply, “when my project is complete, how can you help me digitally publish my results, to meet my funding requirements?” In contrast, recent years have begun to recognize this shortcoming and new university departments have sprung up which focus on providing digital support for humanities, including HRI Digital at the University of Sheffield,⁸⁸ the Köln Center for eHumanities,⁸⁹ the Göttingen Center for Digital Humanities,⁹⁰ and Huygens ING serving the academy across the Netherlands.⁹¹

⁸⁷ Akademie der Wissenschaften zu Göttingen, “Göttingen Septuaginta-Unternehmen,” ADW, 2017, <https://adw-goe.de/en/research/completed-research-projects/akademienprogramm/septuaginta-unternehmen/>.

⁸⁸ University of Sheffield, “DHI - The Digital Humanities Institute, Sheffield,” DHI, 2017, <https://www.dhi.ac.uk/>.

⁸⁹ Cologne Center for eHumanities, “CCeH - Cologne Center for eHumanities,” CCeH, 2017, <http://cceh.uni-koeln.de>.

⁹⁰ Göttingen Centre For Digital Humanities, “GCDH : English,” GCDH, 2017, <http://www.gcdh.de/en/>.

⁹¹ The Huygens Institute for the History of the Netherlands, “Huygens Instituut voor Nederlandse Geschiedenis,” Huygens ING, <https://www.huygens.knaw.nl>.

2. Community

Central to this project is the goal to build an online collaborative research community sharing in work together to produce a community-maintained research dataset which can be used for any purpose. Attaining critical mass for a community is often elusive. This chapter shares the experiences which prompted the design for the components of this research project which are specifically focused on harboring a community of volunteers. An aspect, mentioned prominently in the following section, for retaining a collaborative community is the recognition of work contributed. What would a research paper with a focus on community collaboration be if itself neglected to recognize the community which collaborated to bring it to fruition. The short history that follows, both explains the value of this aspect and other facets of building a thriving online community of collaborators, and also serves to recognize those who have been instrumental to the completion of this research project. This history includes the story of a project which offered lessons that greatly influenced the design for the community of this project.

2.1. History

CrossWire Bible Society started as The White Rabbit BBS (a dialup bulletin board service) in 1984 running self-designed software on a Commodore 64 with two 1541 floppy disk drives. Custom Bible software took advantage of the random access

capabilities of the 1541 drive to quickly lookup any verse of the New Testament-- or at least of the New Testament that had been keyed into the system at the time. IBM's PCs were quickly becoming commonplace and a collaborative project formed between Larry Pierce, a software developer from Canada who created a PC Bible study application called the Online Bible⁹² written in Borland's Delphi programming language; Jerry Kingery, the founder of the Bible Foundation⁹³-- a Bible distribution network based in Oregon; Mark Fuller,⁹⁴ who, with a small group of volunteers, keyed in and checked a number of Bible texts, commentaries, and study resources; and Jerry Hastings, a PC bulletin board system operator⁹⁵ in Phoenix, Arizona who first began online distribution of this free Bible software bundle along with this initially small library of texts. The Online Bible software program was convenient, fast, very useful, and free as in beer, but not free as in speech,⁹⁶ and while the White Rabbit BBS community shifted its efforts from developing their own software, to instead support the efforts of the Bible Foundation, the desire to share in the development effort to improve the software was never met with a welcome. Open source software, with its roots in the Free Software Foundation⁹⁷ and Richard Stallman's frustration over not

⁹² Larry Pierce, "Online Bible - Official Website - North America," Online Bible, 2017, <https://onlinebible.net>.

⁹³ Jerry Kingery, "Bible Foundation - Bible Ministry Resources," The Bible Foundation, 2017, <http://www.bf.org>.

⁹⁴ Some interesting comments from Mark Fuller, along with his home address and phone number at the time-- a sign of how times have changed-- can be seen from the "About" section of one of these early efforts. CrossWire Bible Society, "Webster Bible," The SWORD Project, 2017, <https://www.crosswire.org/sword/modules/ModInfo.jsp?modName=Webster>.

⁹⁵ See "Bible Foundation", Jason Scott, "602 Area Code BBSes Through History," The TEXTFILES.COM Historical BBS List, 2017, <http://bbslist.textfiles.com/602/>.

⁹⁶ For a good explanation of the difference, see in GNU's article, "'free software' is a matter of liberty, not price": Free Software Foundation, "What is free software?," GNU Operating System, 2017, <https://www.gnu.org/philosophy/free-sw.en.html>.

⁹⁷ Free Software Foundation, "Free Software Foundation - working together for free software," FSF, 2017, <http://www.fsf.org>.

having access to the source code which would have allowed him to modify his printer driver at the MIT Artificial Intelligence lab,⁹⁸ was quickly becoming a popular objection to the sweeping wave of protectionism commercial companies were beginning to show to their source code. This protectionism extended beyond commercial companies, to personal developers wishing to keep their source code private for various reason. This included the Online Bible software program, which at the time was only usable on Microsoft DOS. Larry Pierce gave a number of reasons for not wishing to open the source for his program, including financial support it brought through donations to charity organizations which he sponsored. It is not the intent here to evaluate the motives behind the decision, only that keeping the source code closed did not allow a community of volunteers to contribute toward expanding the reach of the software. In an effort to continue to support the cause of the Bible Foundation, the CrossWire Bible Society was officially formed in 1994 around an open source project to provide a cross-platform software system for Bible research. The goal was to bring the functionality of the Online Bible to UNIX and Linux systems, Mac, mobile, and any other platform that might arise. The idea was for CrossWire Bible Society to establish an online community where software engineers could gather in their free time to contribute to open source projects doing the digital work of a traditional Bible Society--distribution of Bible text to a target community. CrossWire's target community was the digital space.⁹⁹ The SWORD Project was and remains the core development effort at

⁹⁸ Sam Williams, *Free as in Freedom: Richard Stallman's Crusade for Free Software* (Sebastopol, CA: O'Reilly, 2002).

⁹⁹ CrossWire Bible Society, "About the CrossWire Bible Society," CrossWire Bible Society, 2017, <http://crosswire.org/about/>.

CrossWire and offers a rich cross-platform library of software components written in the C++ and Java programming languages, and user applications which run on all conceivable devices. In August of 2017, an automatic query¹⁰⁰ of the CrossWire text library for Bibles, commentaries, and other published research material which can be used with the SWORD Project software returns 1,447 items in 797 languages. The online community of developers, testers, text library maintainers, and other contributors numbers at 463 active members.¹⁰¹ The CrossWire Bible Society has partnered with Wycliffe Bible Translators, the American Bible Society, The Society of Biblical Literature, the United Bible Societies, and other translation and ministry organizations on software development¹⁰² and distribution efforts, and on key Bible encoding standards including the Open Scriptural Information Standard (OSIS)¹⁰³ used for many years by Bible societies worldwide to markup and exchange data. The development of this standard contributed important methodologies to the use of XML in general, including the concept of trojan milestones.¹⁰⁴

¹⁰⁰ Using the SWORD Project's remote book installation tool: `installmgr -init -sc -s|grep ^\\|cut -f2 -d|cut -f1 -d|while read -r i; do installmgr -r "$i"; done; echo === Totals ===; installmgr -s|grep ^\\|cut -f2 -d|cut -f1 -d|while read -r i; do installmgr -rl "$i"; done|grep ^\\.\\|wc; echo === Totals by Language ===; installmgr -s|grep ^\\|cut -f2 -d|cut -f1 -d|while read -r i; do installmgr -rl "$i"|cut -d| -f2|cut -d| -f1|while read -r j; do installmgr -rdesc "$i" "$j"|grep ^\\|Lang\\|; done; done|sort|uniq -c|sort -bgr`

¹⁰¹ CrossWire Bible Society, "sword-devel Mailing List Administration, Membership Management, Member List," sword-devel Mailman listserv, <http://crosswire.org/mailman/admin/sword-devel/members>, (URL behind login).

¹⁰² CrossWire Bible Society, "About The Bible Tool," The Bible Tool, 2017, <http://crosswire.org/study>.

¹⁰³ Society of Biblical Literature, "Bible Technology Group Release OSIS 1.0," SBL Forum, 2005, <https://www.sbl-site.org/publications/article.aspx?ArticleId=62>.

¹⁰⁴ Steven DeRose, "Markup Overlap: A Review and a Horse," in *Proceedings of the Extreme Markup* (Extreme Markup Languages, Montréal, Québec, 2004).

2.1.1. Crisis turned catalyst

An unfortunate event occurred in 2003, when Larry Pierce, author of the Online Bible software package, revoked public usage permission for a valuable dataset he had produced which aligned an edition of the Textus Receptus to the corresponding English Authorised (or King James) Version of the New Testament. This event prompted a special team of volunteer software engineers at CrossWire Bible Society to build an online, community collaboration environment facilitating volunteer scholars to produce a comparable replacement, released under an open usage license.¹⁰⁵ The tools for this rushed effort were produced in urgency to meet a gaping hole left by the revocation of the Online Bible software data. This time of crisis quickly brought together a community of developers and researchers with Greek and English skills to complete a sizeable goal within a relatively short amount of time. Important aspects of this project framework, which matured over the life of this effort, developed to address the collaboration needs for this community. The tools built for the project were never intended to live past their initial purpose-- to facilitate the production of the data, though the experience has played heavily in the community collaboration features included in the software system developed for this current research project. The KJV2003 status page, pictured below, allowed an individual volunteer to see, at a glance, a status overview for the entire project, the work to be done, what work was already claimed by another volunteer, what work was claimed by the current user, what

¹⁰⁵ CrossWire Bible Society, "Project KJV2003," KJV2003, <http://crosswire.org/sword/kjv2003>.

work was completed, what work was available and open for a user to claim.¹⁰⁶

Book Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Matthew	12	17	17	17	18	18	18	18	18	18	2	2	2	2	18	18	18	18	18	18	18	18	18	18	18	18	9	9
Mark	20	20	20	20	18	17	18	14	14	18	17	17	17	17	17	17												
Luke	10	10	10	10	10	10	10	10	18	18	18	18	17	17	17	17	17	17	17	17	17	17	17	17	17			
John	16	18	18	18	18	9	9	9	9	3	3	3	23	23	17	17	17	17	17	17	17	17						
Acts	22	22	22	22	22	22	22	17	17	17	17	17	6	6	17	17	17	17	17	17	17	17	17	17	17	17	17	17
Romans	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8												
I Corinthians	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17												
II Corinthians	17	17	17	17	17	17	17	17	17	17	17	17	17	17														
Galatians	23	23	23	23	23	23																						
Ephesians	23	23	23	23	23	23																						
Philippians	4	4	4	4																								
Colossians	21	21	21	21																								
I Thessalonians	21	21	21	21	21																							
II Thessalonians	21	21	21																									
I Timothy	21	21	21	21	21	21																						
II Timothy	21	21	21	21																								
Titus	21	21	21																									
Philemon	21																											
Hebrews	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4													
James	17	17	17	17	17																							
I Peter	6	6	6	6	6																							
II Peter	8	8	8																									
I John	15	15	6	6	6																							
II John	22																											
III John	13																											
Jude	7																											
Revelation of John	7	7	7	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17

■ - Finished 260 / 260 (100%)
■ - Incomplete 0 / 260 (0%)
■ - Unassigned 0 / 260 (0%)

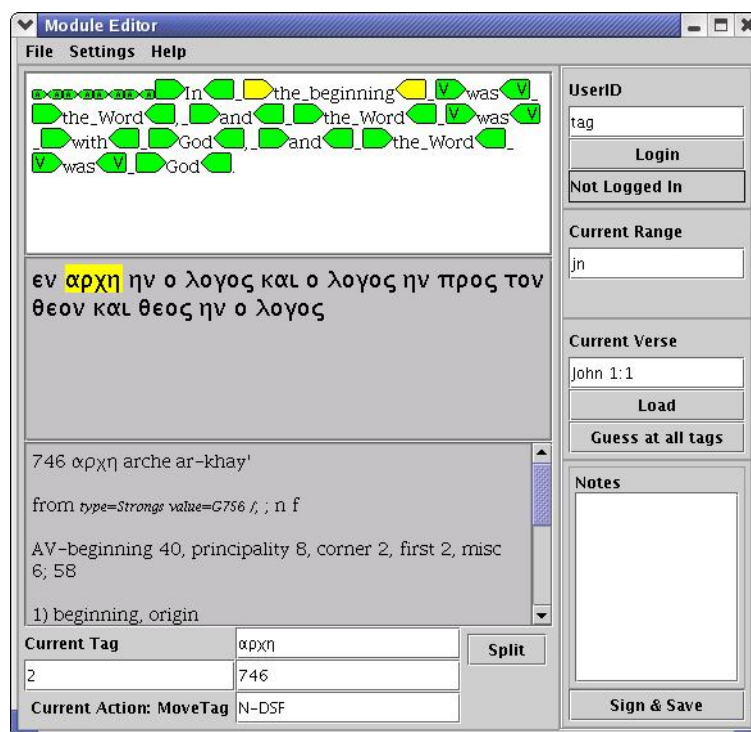
Status page showing the project at 100% completion¹⁰⁷

Each cell represents an individual chapter of the New Testament and also one task of work which could be assigned. The color of the cell designates the status of the task and the number within the cell identifies the volunteer assigned the work. A need was obvious early on for a volunteer to easily communicate with other volunteers, to ask questions and to offer assistance as they worked together on the project, immersing the

¹⁰⁶ CrossWire Bible Society, "KJV2003 - Status," KJV2003, 2017, <http://crosswire.org/sword/kjv2003/status/>.

¹⁰⁷ To see how this tool evolved, see [section 4.4.3. History of indexing tools in the VMR CRE](#).

contributor into a community of fellow laborers. A login mechanism allowed identification and attribution of the work done by each volunteer, giving the scholar a sense of ownership to their contributions. The tool built for performing the work (pictured below) was honed over time to support the volunteer in completing their task efficiently with as much computer assistance as practical, and short of obstructing their workflow.



KJV2003, Contribution Tool

Basic computer assistance used a Greek/English concordance to “Guess at all tags,” initiated with the button of that name, seen above. The volunteer could then click on a Greek word in the Textus Receptus window and adjust the location and width of the associated English tag in the top window. A tag could be split to facilitate

non-contiguous English text associated with a single Greek word, which was an improvement over the data previously used from the Online Bible. While these tools were very crude, the project successfully completed its goal in 2004 with an overwhelming response from volunteers and praise from the same for the opportunity to participate in the work.¹⁰⁸

2.1.2. Garnering associates and interest

In 2005, spurred on by regular solicitation from the CrossWire community for similar collaboration efforts and seeking to incorporate the lessons learned over the span of the KJV2003 project, work began to construct a maintainable, generalized framework of tools which would facilitate future community projects. The desire for one dataset in particular was seeded in the mind of this author during his first international trip to Rome, sponsored by the American Bible Society and hosted by Pope John Paul II in April of 2002 for the release of the newly completed OSIS 1.0 standard to offer training sessions to Bible Societies from around the world. During this trip to Europe, CrossWire held its first international SWORD Meet¹⁰⁹ in Cambridge, UK, just before the OSIS conference, and one of the excursions there was to visit the Tyndale House Research Library where the CrossWire team was invited by the then director Bruce Winter into the “board room” to brainstorm on ideas for further Bible research software

¹⁰⁸ This author remembers emails from a vicar of a small southern Church of England parish who had not used his Greek for years and expressed his joy to have the opportunity to contribute during the week to a project he felt worthwhile. He could not get new assignments fast enough and would often start before we could mark the tasks as reserved. It was a joy to allow others to contribute and hear their stories.

¹⁰⁹ CrossWire Bible Society, *SWORD Meet*, Photo (Cambridge, UK: 2002), http://crosswire.org/~scribe/pics/photo.jsp?l=/2002/0409_rome/2002-04-19_austria_to_cambridge/&i=pict0020_sword_group3.jpg.

development. The meeting included Tyndale's technical director, David Instone-Brewer, who was keen to have more tools available for New Testament manuscript research and offered many ideas from his correspondence with the Tyndale staff. It was a surprise to this author to learn that no digitized dataset focussing on the witnesses of the New Testament was generally available for computer research. Three years past, and while communication remained open between CrossWire and the researchers at Tyndale House, no project materialized until 2005. Interest came from the University of Cambridge via the late Professor Graham Stanton in February of 2005 with an invitation to attend the senior NT seminar on Tuesday, the 22nd at 2:30pm in the Lightfoot Room at the Faculty of Divinity and then afterward share ideas. The meeting produced excitement from both sides and a further introduction was made to Douglas de Lacey, whose skills bridged across both Computer Science and linguistic studies, along with Dirk Jongkind and Peter Head, both maintaining a list of manuscript witnesses¹¹⁰ on Head's personal page on Tyndale's website. This year (2005) also saw the first contact with the University of Birmingham, at the Birmingham SBL annual breakfast, prompted by an invitation from Charlotte Hempel, whose contact was made through a summer course at Tyndale House.¹¹¹ Peter Head and Dirk Jongkind, whose work on the scribes of Codex Sinaiticus¹¹² provided insightful suggestions for tools with the potential to enrich manuscript research; both offered suggestions for

¹¹⁰ Peter M. Head, "Early Greek Bible Manuscripts Project," Tyndale House, 2003, <http://www.tyndale.cam.ac.uk/Tyndale/staff/Head/EGBMP.htm>.

¹¹¹ Tyndale House, Early Christian Backgrounds Summer School 2005, Photo (Cambridge, UK, 2005), http://crosswire.org/~scribe/pics/photo.jsp?l=/2005/SummerSchool/20050729_tyndale_house/&i=imgp1544_scrolls_class_Charlotte_Hempel.jpg.

¹¹² Dirk Jongkind, *Scribal Habits of Codex Sinaiticus* (Piscataway, NJ: Gorgias Press, 2007).

facilitating a collaborative online community of New Testament manuscript scholars. The first development of code for image viewing and transcription, along with a letter back to Graham Stanton was sent in 2006 proposing “to develop a community where scholars can come and register papyri, inscriptions, and the like, upload images, transcribe, translate, annotate, etc. [The] vision is a growing, open community collaborating together to produce free and open data usable by other researchers for any purpose.”¹¹³ Tyndale offered to support the work with a one year scholarship which included residence fees and sponsorship to the UK, to begin in May of 2008 under supervision of Peter Williams, Tyndale’s new warden in replacement of the retired Bruce Winter.

2.1.3. The trek to Münster

Of paramount importance was the introduction at SBL 2008, via Peter Williams and Dirk Jongkind, to Ulrich Schmid-- the maintainer of the *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments* at Aland’s *Institut für Neutestamentliche Textforschung*. Schmid was like-minded about digital technology for manuscript research and had already submitted a project proposal to the Deutsche Forschungsgemeinschaft to begin digitizing and making available online the Institute’s manuscript catalog and comprehensive collection of manuscript microfilm, which first went live in October, 2009. Schmid conveyed the meeting and conversation to Peter Robinson and David Parker at the University of Birmingham and a first proposal for a

¹¹³ Graham Stanton, “Hopeful Research Student,” email, 2006, <http://crosswire.org/~scribe/stanton2006.pdf>.

collaborative online research environment was offered as a proposal for a PhD research project at the Institute for Textual Scholarship and Electronic Editing (ITSEE) in 2008.

¹¹⁴ Per Robinson, ITSEE, at the time was “in the early stages of building a whole new set of tools built on web 2.0 technologies which, we think, are going to shape scholarly editing for years to come.”¹¹⁵ Robinson and Parker offered encouragement over the next months, and at a pub in New Orleans, during SBL 2009, plans were made to collaborate with ITSEE. The match seemed perfect. The next year, in 2010, ITSEE was awarded funding for The Workspace for Collaborative Editing.¹¹⁶ This research and development transitioned after the sponsored year at Tyndale House in 2008, to ITSEE in 2010.¹¹⁷ Peter Robinson relocated to the College of Arts and Sciences at the University of Saskatoon¹¹⁸ to continue his vision building what would become Textual Communities¹¹⁹ though he continued to collaborate and give support to this effort. David Parker and Hugh Houghton took on supervision of this work, offering generous patience, along with ITSEE’s Rachel Kevern and Bruce Morrill, teaching scholarly editing to a software engineer with a Divinity Masters and no previous background in editing. Still, good progress was made on software development alongside the staff at ITSEE and with strategic planning specifically with Houghton. Kevern provided invaluable insights into the needs of a manager directing a group of scholars to do

¹¹⁴ Troy A. Griffiths, “PhD Proposal,” 2008, <http://crosswire.org/~scribe/proposal-bham.pdf>.

¹¹⁵ See early correspondence in the email thread, Peter Robinson and David Parker, “a PhD,” email 2008, <http://crosswire.org/~scribe/robinsonparker2008.pdf>.

¹¹⁶ ITSEE, “The Workspace for Collaborative Editing,” University of Birmingham, 2013, <http://www.birmingham.ac.uk/research/activity/itsee/projects/workspace.aspx>.

¹¹⁷ UK Visa issues occupied 2009.

¹¹⁸ Peter Robinson, “Staff Profile: Peter Robinson,” University of Saskatchewan College of Arts and Science, 2015, <http://artsandscience.usask.ca/profile/PRobinson>.

¹¹⁹ Peter Robinson, “Textual Communities,” University of Saskatchewan, 2017, <http://www.textualcommunities.usask.ca>.

indexing and transcription work. The project management and status components mentioned in the next sections were extended to accommodate her feedback.

In July of 2011, Martin Fassnacht, who led the technology efforts at the INTF developing the VMR 1.0 software resigned his role on the 3 year *Deutsche Forschungsgemeinschaft*-funded project to take up a permanent post with the University of Tübingen. At the recommendation from Houghton to the INTF, this author was presented with the opportunity to assume the vacated role for the remaining 2 years of the project. This development was a catalyst to the acceleration and adoption of the community collaboration framework by the INTF.

Initially, there began a period of building rapport with the staff at the Institute by gently offering a genuine look and honest evaluation of the progress of the project over the first year. Many historical milestones had been crossed and valuable data produced. The Institute needed to see that sincere value was ascribed to their work thus far. There were many opportunities for improvement and time afforded a perfect occasion to plant the seed of excitement about the possibility of a truly open environment where collaborative research and the effort to build and enrich the data would go hand in hand. Many of the staff at the INTF launched head first together into the work, including Schmid. Working daily alongside a true domain expert and potential user of the software able to shape the vision into what would be a welcomed tool from the scholarly community was invaluable. Marie-Luise Lakmann, who among many other things, manages transcribers for the INTF ECM work, along with her students, became

the Quality Assurance team, testing software in live work, long before it was stable.

Georg Gabel became my User Interface test subject, trying new designs for regularization editing and giving feedback on their usefulness. Christian Askeland, the Coptic expert employed by the *Apokalypse* ECM Project in Wuppertal, constantly pushed for new functionality to assist his research and willingly assumed the role of Quality Assurance Tester, finding issues and giving feedback before general availability. Klaus Wachtel along with Holger Strutwolf kept the ship afloat and on course while Schmid helped this research project remove and replace bow planks that invariably obstructed the work on Acts-- with the hope that Mark, Matthew, and the Apocalypse would be completed in record time using the new software. Catherine J. Smith partnered as a fellow Software Engineer to contribute ITSEE's Collation and Apparatus Editor, collaborating on a seamless integration into the software system. Martin Sievers and his team at the University of Trier also contributed a key component, the Online Transcription Editor. The Interedition regulars, including Tara Andrews, Ronald Dekker, Gregor Middell, and Joris van Zundert, contributed endless advice and without their work on CollateX, the realtime collation aspects which make this project functional would not have been possible. The invaluable conversations during three years of daily work next to the researchers at the INTF, including Schmid, Wachtel, Gabel, and Lakmann, are unfortunately impossible to list in a bibliography.

Everyone mentioned in this short history has been instrumental to bringing this research project and resultant software system to completion. The contact with Schmid initiated in November 2008, would solidify into the core collaboration, and friendship, which

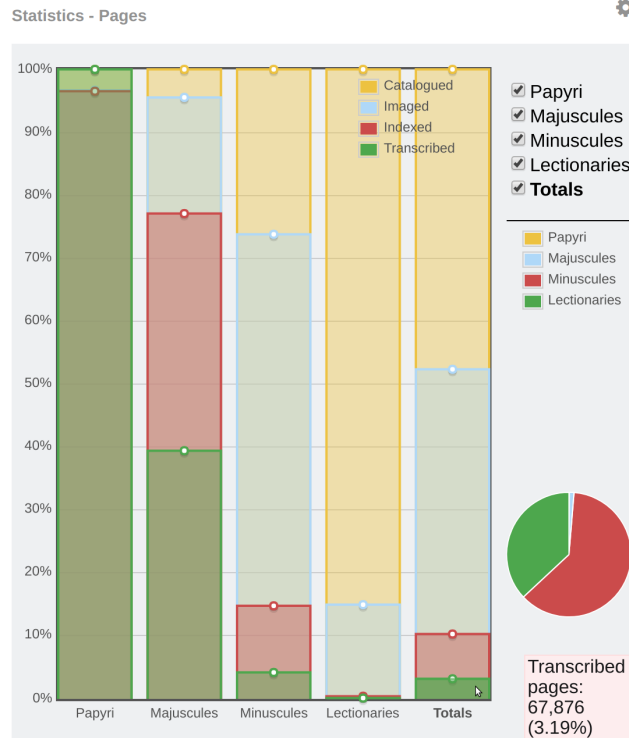
has produced the VMR CRE-- which, along with the contributed components from the Workspace for Collaborative Editing, has brought the editing workflow at the INTF for the ECM into a collaborative online platform-- “a community where scholars can come and ... transcribe, translate, annotate... a growing, open community collaborating together to produce free and open data usable by other researchers for any purpose.”

This history reflects the importance of community in academic software development, where filling in the gaps to gather a full software team must be intentional. The latter half of this history is meant as an encouragement and admonition to the lone humanities project software developer. Shared below are lessons which can be drawn from the first half of this history, serving as a software project example which successfully built an online collaborative community.

2.2. A project with a goal

A project can expect a first-time visitor to their website to spend a very short time before navigating away. The experts say a site has about “10–20 seconds, but pages with a clear value proposition can hold people's attention for much longer. To gain several minutes of user attention, you must clearly communicate your value proposition within 10 seconds.”¹²⁰ The first step to draw an able, potential participant into an online community, is to immediately make a visitor understand what the project is attempting to accomplish and see that it is heading toward that goal.

¹²⁰ Jakob Nielsen, “How Long Do Users Stay on Web Pages?,” Nielsen Norman Group, September 12, 2011, <https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages/>.



NTVMR Statistics

2.3. A project with life

Next, a would-be volunteer needs to know they are not wasting their time on a project which is dead or dying. The first page of a project should make evident that life is present-- that people are actually sharing in work together to reach the goal. The most recent contributions and activities should be clearly seen on the first page of a community collaboration effort. Online users, activity streams, and recent forum conversation summaries are great ways to convey that the project is alive. Community is inherently attractive to human beings. Displaying prominently the activity of a

thriving community capitalizes on the rare commodity of momentum, assuring a collaborative project continues to grow.

Recent Additions

Latest additions to the New Testament Virtual Manuscript Room. Last modification date given for each document. Numbers reflect activity for the past 10 days.

1571 (31571)

8/27/17 7:17 AM

43 verses indexed by irenaeus

356 (30356)

8/27/17 4:18 AM

272 verses indexed by gmbarnhill

138 (30138)

8/27/17 2:26 AM

85 verses indexed by bruce

17 features annotated by bruce

1829 (31829)

8/26/17 3:37 PM

355 verses indexed by agp

105 (30105)

8/26/17 10:50 AM

231 verses indexed by despoina

46 (30046)

Activities

Wrote a new forum post in External Images.

GA 86

Colour images of GA 86 are available at <http://digitalna.kniznica.info/zoom/68240/view?page=1>

Wrote a new forum post in Page 920 (folio 44v).

31739 (1739) Page 920 (folio 44v) General Discussion

I have only checked a few folios but the transcriptions for 1739 seem to be incorrectly matched to their corresponding folios.

Replied to Peter J Gurry's forum post in Conjecture ID ci10092


RE: J. H. Ropes supported the conjecture

Thanks for mentioning this. It will be included in t of the additions to be uploaded around 1 October & Cadbury (1933),...

Activity Streams, Recent Forum Posts, and Online Users

2.4. Easy level of entry

Once convinced that a project is worth their precious time, a new contributor needs to know how to start. No obstacles should remain which keep the volunteer from claiming responsibility for, and beginning an entry level task. Some work on a project will always require advanced skill and supervision, but there should be at least one task that a new volunteer can freely accept and begin without approval. The result does not always have to be made public before review, but their acceptance of responsibility should immediately be shown.

Indexing Status 

Pending Work

The INTF NTVMR has complete Biblical Content index information for hundreds of New Testament manuscripts. Below you will see a list of manuscripts which have not yet been fully indexed but have images available. You can help. Become an active member of the NTVMR community and help us build a more complete manuscript repository for NT studies.

Status	Indexer	INTF Id	GA	Date	Pgs
Results: 2323					
66% Complete (View)	sheekh3arb	20002	02	V	144
99% Complete (View)	lesjanc	20019	019	VIII	257
89% Complete (View)	ehixson	20023	023	VI	44
12% Complete (View)	mnims	20027	027	VII	48
25% Complete (View)	tagriffitts	20029	029(+0113+0125+0139)	V	32
90% Complete (View)	Claim Responsibility	20038	038	IX	249
4% Complete (View)	Claim Responsibility	20040	040	VI	89
96% Complete (View)	drubschmid	20042	042	VI	188
41% Complete (View)	Claim Responsibility	20043	043	VI	197
9% Complete (View)	drubschmid	20047	047	VIII	152
29% Complete (View)	Claim Responsibility	20055	055	XI	303
15% Complete (View)	Claim Responsibility	20056	056	X	381
50% Complete (View)	Claim Responsibility	20057	057	IV/ V	Frg
33% Complete (View)	Claim Responsibility	20067	067	V	2

NTVMR Indexing Task List¹²¹

¹²¹ To see how this tool evolved, see [section 4.4.3](#).

Once a volunteer has taken that first step to claim responsibility for a piece of work, they then should visibly see their name identified with the project. This sense of identity brings many advantages for a project. First, the volunteer feels a sense of belonging which encourages the individual to remain involved. Second, there is established a joint responsibility for the success of the project among its members. Next, an embarrassment for long outstanding assignments, along with a sense of pride for completed work is displayed for all to see. And finally, a no-loss/win scenario is created by the element of community collaboration; i.e., while asking for help is by no means seen as a negative to a project, when a volunteer benevolently contributes to the same, he or she is the object of praise. This encourages the recipient of the praise to advertise the project. As Adam Smith rightly observes, “Though the meer want of beneficence seems to merit no punishment from equals, the greater exertions of that virtue appear to deserve the highest reward. By being productive of the greatest good, they are the natural and approved objects of the liveliest gratitude.”¹²²

¹²² Adam Smith, *The Theory of Moral Sentiments* (London: A. Millar, 1761), 138.

2.5. A project with a soul

During the efforts for the KJV2003 project¹²³ it became clear early on that the volunteers all had the same questions and also that many of them were happy to offer solution to problems they faced while completing their assigned tasks. A crude, “Email all volunteers” link allowed the team to easily shout-out to others when any had a question. Online collaboration has progressed since 2003 but the principle remains the same. Community is not community if people cannot speak with each other. Forums, chat, blogs, mailing lists are all vital ways to help bring humanity to your digital humanities project. Raising the ease of communication within the system and the visibility of that communication between members is the objective. Opportunities for real collaboration begin with communication. The occasion for questions to be asked and answered, while necessary and useful, is subordinate in value to the sense of community these opportunities afford. Communication opportunities are one of the most important aspects for building a sustained and thriving online collaborative community.

2.6. Attribution / credit collaborators

A scholar’s reputation is a key facet to the value of their contribution to a project, and attribution in the academic community is an historical central foundation for scholarship, without which, it would be impossible to evaluate reputation. Contributed

¹²³ See [2.1.1. Crisis turned catalyst](#), above.

work in an online collaborative research environment must maintain this traditional element of attribution. If a contributor knows they can recall all of their work easily, and thus others can do the same, they retain a traditional sense of amassing their own portfolio of research, and are governed by the peer review concept which has well driven the academic institute for centuries. The task assignment and progress page mentioned earlier is a great way to let others know who is working on a task, and also the activity stream raises visibility for active volunteers. But retaining the contributor's identification with the contributed data is worth more than simply feeding a contributor's pride. This credit, and the ability to summarily recall this credit, is valuable in many ways to a volunteer scholar. Keeping them aware the work they are submitting will continue to attribute credit to the author is very important in academia.

2.7. Natural contribution

Finally, critical to the success of any software adoption is the value of that software to enhance the research process, rather than obstruct an efficient workflow already in place. New software should solve problems which exist in the current environment to prompt adoption and loosen resistance to change. Build a tool which meets a need, fills a gaping hole that you hear a scholar complain about.

Now, once this research tool begins to attract consumers, effort should be made to give them tools to add data naturally while they use the system for their own research. For example, if index data has not yet been entered for a set of manuscript images, the images still provide value to the scholar, even if they will be required to spend the time

to find the page which interests them. At this point, a utility should be available for the scholar to add the index data to that page. They have done the work to identify text within the image set. Clearly present a button in the user interface, "Index this page." The scholar has the desire to add the data at this point. Provide them an opportunity to contribute to the project, which will not only improve their experience the next time they need to locate the same page, but also the experience for the rest of the community.

3. Design

Effective software *development* begins with the identification of those individuals who will fill the established software development roles described in [Section 1.4](#).

Successful software *design* begins with the identification of the roles, or actors, who will interact with the software system and the ways in which they desire to interact-- their **use cases**. From Booch,¹²⁴ “a use case represents a *complete* flow of activity of *what* the actor wants your system to do from the *actor’s point of view* that provides *value* to the actor.”¹²⁵ Determining actors is not the same as determining the audience for a software system, though this can help. When one thinks of the user base for a system like the New Testament Virtual Manuscript Room, what comes to mind is a list like: Textual Scholar, University Professor, Seminary Student. While the target audience is necessary to identify, they are not primarily what help us build our use cases-- they are not specific enough to describe their interaction with the software system to serve as an actor in use case compilation. The actor title incorporates their purpose for and role in a specific use case. For example a “transcriber” may wish to save privately their transcription work still in-progress before they are ready for their work to be seen by anyone else. This scenario is specific, concise, and is an example of a valid and valuable use case to include in the design for this work.

¹²⁴ Grady Booch is one of the fathers of UML design, which proposed the concept of 'use case'.

¹²⁵ Grady Booch et al., *Object-Oriented Analysis and Design with Applications* (Boston, MA: Pearson Education, 2007), 262.

After relocating to Münster in June of 2011 to begin this work on what has now become the Virtual Manuscript Room Collaborative Research Environment (VMR CRE),¹²⁶ Ulrich Schmid was assigned as my officemate and collaborator on the work. Schmid is a long recognized expert in the field of New Testament studies and textual criticism, had worked at the INTF for a number of years, providing intimate knowledge of the ECM workflow, was 50% of the 2 person team imagining and developing the VMR 1.0, and was the curator of the now online *Kurzgefaßte Liste*. Schmid was immediately identified as **business analyst** to liaise between the expert user and the software development process. The next task, with Schmid's help was to fully understand and critically evaluate the work done to this point on the VMR 1.0. The evaluation which follows highlights the advantages for applying a classical computer science design and development model by reviewing a project begun by well-intentioned humanities experts without the assistance of a computer science specialist. In addition, this review will contribute to the design process by evaluating the use cases which were serviced by the feature set included in the VMR 1.0.

3.1. VMR 1.0: a critical evaluation

In short, the Virtual Manuscript Room 1.0 implementation embeds all business logic,¹²⁷ SQL statements, and web form navigation logic into PHP pages to access data from roughly 450 MySQL database tables-- one separate database table reserved for each

¹²⁶ Troy A. Griffiths, "Virtual Manuscript Room Collaborative Research Environment," VMR CRE, 2018, <http://vmrcr.org>.

¹²⁷ In computer science, 'business logic' refers to the layer of functionality which captures the design model and domain logic, typically implemented just above the datastore layer and below a user interface.

manuscript in the catalog. There is no formal definition of a **data model** to give cohesion to architecture design. There is no **tiered design** consideration to isolate and capture the intricacies of **business logic** and promote reuse. There is no evidence of forethought toward a **service oriented architecture** to enable integration with other institutions. There are no **normalization** principles applied to the database design, limiting the types of questions which can be asked of the data. The majority of the data captured has not been **regularized**, making it useless for digital analysis. No **coding style** or naming conventions have been declared or internally followed, making the code and database unintelligible and unmaintainable-- beyond its original author-- within a team of developers, or any future developer who has not been made privy to the hidden nuances therein. No source code **revision control** system has been implemented, losing all traceability of the reasoning behind the current state of the software. The benefits of a classically designed infrastructure far outweigh the successful delivery of a finite set of features to end users as a single product at a temporal point in technology history. The discipline of computer science grants **longevity** to a system by engineering a design which is optimized to

- **convey its design** to other engineers-- within the design implementation itself, within clear and concise design documents, and through **traceability** within the revision history of a source repository,
- **prevent** the possibility of **inconsistencies** in data by following unanimously accepted 3NF data normalization principles¹²⁸

¹²⁸ See Codd, 3NF in "Normalized Data Base Structure: A Brief Tutorial," in *Proceedings of the 1971 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control*, SIGFIDET '71 (New York, NY: ACM, 1971), 1–17, doi:10.1145/1734714.1734716.

- encapsulate and isolate logic into **reusable** components so other parts of the system need not duplicate logic and future components can build upon the work done, and
- enable **maintainability** by ensuring components do **discrete** work, promoting easy isolation and repair of problems.

When systems are not built upon this solid foundation of classical computer science design principles which have enabled systems to perpetuate for decades, a system quickly becomes inextensible, unintelligible to other engineers, and inevitably is rebuilt again and again losing all expert knowledge captured in any previous incarnation.

3.1.1. Frontend / user interface

The Virtual Manuscript Room 1.0 is implemented as a web application having 3 major facilities (for screenshots, see the introductory [Section 1.5.3](#)):

1. Manuscript search facility
2. Manuscript image/transcription viewing facility
3. Manuscript indexing facility

The search facility is fairly limited in its feature set due to the definition of the system's data model (see next section). The only parameters available for searching are all mutually exclusive and consist of:

- Object ID

- Gregory-Aland Number
- Former Gregory-Aland Number
- Country
- Place
- Institution
- All Manuscripts With Images

Again, no combination of these parameters can be supplied, but only a single value of only one parameter. The first 3 options should be classified as simple navigation functions rather than search. The next 3 controls work together to constitute one geographic search of the current location data of the manuscripts, each parameter having a progressively more precise specification. The final option by itself has no particular use except to assess the completeness of the image repository of the Virtual Manuscript Room. In conclusion, there exists a lookup feature, if one knows the manuscript to which she would like to view, and a current location search, which might come in handy if one plans to travel to a region and would like to know which manuscripts are available in the target area.

Next, the viewing facility has two modes: a 'reading' view which shows a manuscript page transcription next to a single image of that page, and 'skyscraper' view which presents an image gallery (in what looks similar to the tall set of parallel windows of a skyscraper) focusing on the task of image viewing. Neither of the viewers allow a URL

to be constructed to point a visitor to a particular image of a particular manuscript. This prohibits any other website from linking to anything but an entire manuscript in the Virtual Manuscript Room 1.0, or further, from any other writing or online resources referencing a specific folio side or transcription.

Finally, the indexing utility, designed to facilitate a privileged user's contribution to the index material within the system, consists of a single form which allows maintenance of a concept which the VMR 1.0 calls 'current numbers'. These 'current number' entities do not adhere to a single coherent usage definition but instead have many personalities. Sometimes a 'current number' represent a manuscript page, another might represent a single source of images for a single page, and another might represent the index information for a single contiguous segment of biblical content on a single page. Sometimes they represent any combination of any of these concepts. All 'current numbers' for a single manuscript are stored in 1 of approximately 450 tables, all with the same data definition, each reserved for a specific manuscript. Basically, this indexing facility simply enables create, read, update, and delete (CRUD)¹²⁹ operations on these tables.

3.1.2. Datastore

Moving on from the frontend and focusing on the underlying data model of The Virtual Manuscript Room 1.0, we begin to see why the user interface does not expose more

¹²⁹ For more on standard CRUD principles within database design, see Ryan Stephens and Ronald Plew, *Database Design* (Carmel, IN: Sams Publishing, 2000), 237ff.

research queries. The first and most glaring liability is shown in the previously mentioned duplication of the same data definition, once for each manuscript-- more than 450 tables, each designated to hold information about the folios for a single witness. This makes querying across all witnesses virtually impossible. This shows that, while the staff have ample experience in the field of manuscript studies, there is a fatal absence of formal training in computer science.

3.1.2.1. Organize by type, not content

Fundamentally, a relational database design must center around the ***type*** of data to be captured, not the variation in content. For 60 years, the INTF has organized their data within traditional filing cabinets, each with boxes of microfilm, or a hanging file-- one for each manuscript, so it is understandable that their initial database design might have mirrored this approach, having one table for each manuscript. But this model is not at all conducive to a useful electronic database design. The filing cabinets and hanging folders center around organizing the various ***contents*** of the data, whereas the INTF electronic database design should have focused on the various ***types*** of data.



Ulrich Schmid and Rachel Kevern, Classical Database

A brief example will help press home the severity of this design flaw. Consider excerpts from 3 tables organized by content:

Manuscript 1

Folio Number	Start Verse	End Verse	Image URL
1r	John.3.16	John.7.9	image1_1r.jpg
1v	John.7.9	John.11.17	image1_1v.jpg

Manuscript 2

Folio Number	Start Verse	End Verse	Image URL
47v	John.3.16	John.5.12	image2_47v.jpg
48r	John.5.12	John.7.14	image2_48r.jpg

Manuscript 3

Folio Number	Start Verse	End Verse	Image URL
5r	John.1.1	John.3.16	image3_5r.jpg
5v	John.3.16	John.4.2	Image3_5v.jpg

Now suppose we would like to find all folios with their images which begin with John.3.16. Our corresponding SQL query would look something like this:

```
SELECT 'Manuscript 1' as Manuscript, `Folio Number`, `Image URL` FROM
`Manuscript 1` WHERE `Start Verse`='John.3.16'

UNION

SELECT 'Manuscript 2' as Manuscript, `Folio Number`, `Image URL` FROM
`Manuscript 2` WHERE `Start Verse`='John.3.16'

UNION

SELECT 'Manuscript 3' as Manuscript, `Folio Number`, `Image URL` FROM
`Manuscript 3` WHERE `Start Verse`='John.3.16';
```

Yielding the result:

Manuscript	Folio Number	Image URL
Manuscript 1	1r	image1_1r.jpg
Manuscript 2	47v	image2_47v.jpg
Manuscript 3	5v	image3_5v.jpg

Imagine now our SQL query if we had not 3 but 5000 manuscripts cataloged.

Instead, recognizing that we wish to store *page type* data, creating a Pages table would take advantage of electronic database functionality. Here the organizational focus changes from “Manuscript 1, 2, 3, ...” to “Pages”-- from content centered, to type centered:

Pages

Manuscript	Folio Number	Start Verse	End Verse	Image URL
Manuscript 1	1r	John.3.16	John.7.9	image1_1r.jpg
Manuscript 1	1v	John.7.9	John.11.17	image1_1v.jpg
Manuscript 2	47v	John.3.16	John.5.12	image2_47v.jpg
Manuscript 2	48r	John.5.12	John.7.14	image2_48r.jpg
Manuscript 3	5r	John.1.1	John.3.16	image3_5r.jpg
Manuscript 3	5v	John.3.16	John.4.2	Image3_5v.jpg

The modification is minimal; all we have done is add one extra column to the previous design, designating to which manuscript the folio belongs. This yields a single table design to replace what would be an ever increasing table count with potentially 5500+

duplicate table definitions, simply by changing focus of the table definition from content-centered to type-centered. Electronic databases are exceptionally quick at finding the content we want. We do not need to help them by organizing the content of our data. Our SQL query to yield the same result, if we have 3 manuscripts or 5000, now simply becomes:

```
SELECT Manuscript, `Folio Number`, `Image URL` FROM Pages WHERE `Start Verse`='John.3.16'
```

This concept of type-centered database design is commonly missed by the untrained and can be seen in most projects at the INTF in how they duplicate database tables, one for each biblical book, or each chapter within a book, etc. This design flaw prohibits useful queries across their entire dataset. The researchers at the INTF commonly browse through their databases as they did their steel filing cabinets and may feel overwhelmed to see all folio data grouped into a single table. This can be alleviated simply by defining database views for specific work cases if the database will be browsed directly. For example,

```
CREATE VIEW Papyri as SELECT * FROM DOCUMENT WHERE DocumentID  
BETWEEN 10001 and 19999;
```

```
CREATE VIEW Acts as SELECT * FROM CBGM WHERE BookNumber = 5;
```

What follows is a detailed review of the individual data structures for the VMR 1.0. As a convention for this chapter, database table names and column names used in prose will be underlined for clarity.

3.1.2.2. HSSListe table

The primary table, HSSListe (defined in full below), violates same-object principles by playing double duty to store both *manuscript-centric* rows and also rows for '*holding institution*'¹³⁰ *shelf number*¹³¹ *information*. To determine which of these roles a row in this table is playing, one needs the special knowledge to look at a secondary key column InstID and apply the convoluted logic which follows: if the value of this column begins with a '0' and the next character is not a '0', or this column value is a single digit '0' then we are playing the role of *holding institution shelf information data*; otherwise, if the integer value of this column resolves to 0 (e.g., '00', '000', etc., but not a single '0'-- already described) then we are playing the part of *manuscript-centric information*. Besides this multiple personality triggered by a magic incantation, the remainder of the 56 columns in this table include normalization deficits listed as follows.

First, information is stored which is derived from other columns. This introduces opportunities for inconsistencies. For example, fields GA, GA prae, GA zahl,

¹³⁰ A holding institution is any library, museum, university or other institution which possesses an ancient artifact.

¹³¹ A library shelf number is simply the identification a library has assigned to the object.

GA_post are intended to store the Gregory-Aland number for the manuscript. The first is the complete GA number, the next is the prefix part of the GA number-- or one might think from the name; it is actually used to store the first digit of the Object ID-- a unique integer given to each manuscript. GA_zahl stores the integer value of the Object ID modulus 10000, and GA_post actually stores any extraneous text after the GA number in the GA field. for example:

ObjID	GA	GA_prae	GA_zahl	GA_post
10011	P11(+14?)	1	11	(+14?)

Not only is the naming of these columns misleading, their purpose is useless and their presence violates a primary tenant of data normalization: avoid opportunity for inconsistency. For example, all of the following represent inconsistencies in the data:

ObjID	GA	GA_prae	GA_zahl	GA_post
10011	P11(+14?)	2	11	(+14?)
10011	P11(+14?)	1	12	(+14?)
10011	P11(+14?)	1	11	(Hi Mom!)

A normalized representation without opportunities for inconsistency might look like this:

ObjID	GA	GA_post
10011	P11	(+14?)

GA_prae can be derived from ObjID divided by 10000, and GA_zahl from ObjID modulus 10000. Argument could also be made to remove GA entirely and derive it from:

```
CONCAT(  
    (CASE(CAST( ObjID/10000) AS SIGNED)  
        WHEN 1 THEN 'P'  
        WHEN 2 THEN '0'  
        WHEN 3 THEN ''  
        WHEN 4 THEN 'I'  
    END),  
    ObjID % 10000)
```

This final approach would leave no room for inconsistencies, though also no room for intended deviation from 10000, 20000, 30000, 40000 numbering categories chosen by the Institute, in the event of an exception, thus a practical compromise would weigh in favor of keeping the GA field.

This same normalization violation exposes itself in the fields:

- copies, exemplar - copies is derivable from exemplar
- StatusImage - can be derived from presence of images
- StatusTrans - can be derived from presence of transcriptions

- StatusIndex- can be derived from presence of index data
- Olim, OlimPrae, OlimZahl, OlimPost- same as GA

The next normalization violation is seen in the pattern to repeat column names suffixed with a number, as an attempt to store, within a single database row, data which might repeat. These columns try to use naming patterns to indicate their repeating logic. Instead, normalized database design relegates the responsibility of repeating data to a table explicitly designated for the purpose. For example, the violating structure is shown from the following table excerpt:

ObjID	Img1	Img2	Img3	catalog	catalog2	catalog3
10048	http://www.accademiafiorentina.it/paplett/big.asp?nome_img=PSI+1165.jpg&sigla=PSI+X+1165	NULL	NULL	http://www.accademiafiorentina.it/paplett/scheda.asp?id=184	http://opac.bmlonline.it/Record.htm?record=853512467179	NULL

Beyond the aforementioned naming inconsistencies (upper/lowercase, abbreviated and unabbreviated), there is also no dependable pattern to denote the repeating field groups, as can be seen by the lack of '1' in the catalog column name compared to Img1.

Functionally, this data model definition unnecessarily limits to 3 instances in each group. Here, these column groups hold URL references to external websites, pointing to either an available image repository for this manuscript, or an external catalog entry for this manuscript. This is not easily determined from the column names. A

normalized representation using an explicit table definition for the purpose to store external repositories can be represented as:

External Repositories

ObjID	RepositoryType	RepositoryURL
10048	IMAGE	http://www.accademiafiorentina.it/paplett/big.asp?nome_imm=PSI+1165.jpg&sigla=PSI+X+1165
10048	CATALOG	http://www.accademiafiorentina.it/paplett/scheda.asp?id=184
10048	CATALOG	http://opac.bmlonline.it/Record.htm?record=853512467179

The benefits of this normalized table design include:

- explicit association of all data from these six columns as being the same kind of data. This design self-describes its function to future developers. There is no guessing or special knowledge needed to implicitly form a logical group from similarly named columns.
- removing the need to hardcode six column names in SQL and six times repeated logic to determine what data is available. A simple query to get all external repositories for this manuscript changes from looking in six different columns, to a simple join to the secondary table asking for all external repositories.
- the possibility to extend repository types beyond IMAGE and CATALOG, seen here, without changing database structure. For example, one could easily add entries for external TRANSCRIPTION or BIBLIOGRAPHY repositories without changing the data model.
- not storing empty columns when there are not six external repositories for a manuscript.
- removing the artificial limit of three image repositories and three catalog repositories. Any number of external repositories can be recorded.

The next violation of classical software design practice is seen in application functionality pushed down into the database tables, with columns in the HSSListe table such as viewer, making impossible the separation of the current VMR 1.0 functionality from the all-important data which will have any number of other future applications.

Isolating the design focus of the datastore to what the data **is**, rather than what the data is used **for**, with regard to the current project at hand, grants the precious data repository longevity, keeping it untangled and uncluttered from values meaningless outside the current software implementation. This will also lead to capturing the **right kind** of data, making the dataset more valuable, e.g., for a particular manuscript, instead of choosing to store a flag **for** whether or not to present the image viewer in the current application, one might rather store what the Distribution License **is**. This field would be beneficial to any number of other applications and still serves the same purpose for the VMR 1.0-- whether or not to present an image viewer to the user. This awareness to separate what data “is” and relegate that concept to the datastore, from what the data will be used “for” in one incarnation of a software system, is a vital skill for designing a database for longevity.

The HSSListe table attempts to enumerate as columns all possible features which a manuscript might possess. This design both limits the features a researcher might wish to record about a manuscript, and also inefficiently stores data when most of the fields for a particular manuscript are left blank, which is more often the case. For example, note the sparse data shown in this excerpt, with one row attempting to capture every possible manuscript feature:

ObjID	Destroyed	Ink	Canvas	Palimpsest	PalemsestOther	VonSodenID	TischendorfID	OxyID	...
20229	true			undertext	Coptic Script				

3.1.2.2.1. Entity Attribute Value (EAV) design

A more efficient and dynamic design would be to provide a secondary Manuscript Features table to capture only those features which apply to each manuscript. For example:

Manuscript Features

ObjID	Feature	Value
20229	DESTROYED	true
20229	PALIMPSEST	undertext
20229	PALIMPSESTOTHER	Coptic Script

This Entity-Attribute-Value (EAV) design¹³² is used frequently in domains which capture small quantities of data, relative to the total number of all possible kinds of data which could be captured. One such domain which uses EAV data model design extensively is that of Health Services in the storage of medical records. At one health checkup, a doctor may record only a handful of statistics about the patient, when the realm of total possible attributes which could be recorded numbers in the millions. EAV structures also allow for an ever-expanding set of attributes to be captured. An EAV design for manuscript features will allow both dynamic expansion to include new attributes of a manuscript without requiring database changes, and will not force each manuscript to store NULL values for all columns which do not apply.

Two drawbacks of this suggested EAV design are first, the mixing of data types in the Value of the entity (e.g., boolean, integer, enumeration, string), and second, a loose definition of what features should be recorded. These deficits have solutions. For example, a Feature Definition table might define each possible feature to be recorded, along with the expected data type. For strictly typing the value data, the Manuscript Features EAV table could include a value column for each data type:

Feature Definition

FEATURE	DESCRIPTION	DATATYPE
DESTROYED	Whether or not a manuscript is destroyed.	BOOLEAN

¹³² Valentin Dinu and Prakash Nadkarni, "Guidelines for the Effective Use of Entity-Attribute-Value Modeling for Biomedical Databases," *International Journal of Medical Informatics* 76, no. 11–12 (2007): 769–79, <https://doi.org/10.1016/j.ijmedinf.2006.09.023>.

PALIMPSEST	Whether or not a manuscript is one text of a palimpsest, and if so, is our document the overtext or the undertext	ENUM [undertext, overtext]
PALIMPSESTOTHER	If a manuscript is one text of a palimpsest, what is the other text	STRING

Manuscript Features

ObjID	Feature	Value Boolean	Value Integer	Value String
20229	DESTROYED	TRUE	NULL	NULL
20229	PALIMPSEST	NULL	NULL	undertext
20229	PALIMPSESTOTHER	NULL	NULL	Coptic Script

While this design allows data type validation, appropriate sorting, and other benefits that come with having clearly typed data, it re-introduces the inefficiency of NULL column values for type fields that do not apply. As the total number of manuscript and page metadata feature definitions grows, while the average number of recorded features per manuscript remains constant, this proposed EAV design will yield more efficient storage, even with the strongly typed value suggestion given above. The primary benefit achieved from this design is a dynamically expandable feature set for the scholar, while still retaining a strict definition of each possible feature.

If it has not yet been clear from the samples already mentioned, the HSSListe table and generally all the tables in the VMR 1.0 database suffer from column name

inconsistencies, using both upper and lower case differently (e.g., GA_prae, OlimPrae), underscores ('_') are used inconsistently (*idem*), German and English are both utilized (e.g., StatusIndex, Inhalt), abbreviations are sometimes used and sometimes not (Jh, Datiert 1). This makes it extraordinarily difficult to remember the names for each column and practically mandates the table definitions be constantly in front of the developers to get the names correct. In general, not only are the column naming patterns inconsistent, the names chosen for the columns often yield no indications to their intended usage (e.g., Dub, Olim, Zer, Anmerkung, Bemerkung, Img1, catalog).

Following is the complete HSSListe table definition:

HSSListe Table

Field	Type	Null	Default
ObjID	mediumint(5)	YES	NULL
InstID	varchar(20)	YES	0
Dub	smallint(1)	YES	0
GA	varchar(20)	NO	NULL
GA_prae	tinyint(4)	YES	NULL
GA_zahl	smallint(10)	YES	NULL
GA_post	varchar(20)	YES	NULL
Olim	varchar(255)	YES	NULL
Inhalt	varchar(500)	NO	NULL
Inhalt2	varchar(10000)	YES	NULL
Sprache	varchar(100)	YES	NULL
Jh	varchar(10)	NO	NULL

Datiert_1	smallint(20)	YES	0
Datiert_2	smallint(20)	YES	0
Beschreibstoff	varchar(20)	NO	NULL
Pal	varchar(100)	YES	NULL
PalB	varchar(500)	YES	NULL
Blattzahl	varchar(100)	NO	NULL
Blattzahl_Int	int(11)	YES	NULL
Textträger	tinyint(10)	YES	0
Spalten	varchar(10)	NO	NULL
Zeilen	varchar(50)	NO	NULL
Zeilen_R	varchar(40)	YES	NULL
Höhe	varchar(50)	YES	NULL
Höhe_R	varchar(40)	YES	NULL
Breite	varchar(50)	YES	NULL
Breite_R	varchar(40)	YES	NULL
Format	varchar(250)	NO	NULL
Aufbewahrungsort	varchar(200)	NO	NULL
Shelf	varchar(200)	YES	NULL
Soden	varchar(20)	YES	NULL
Tischendorf	varchar(20)	YES	NULL
Oxy	varchar(100)	YES	NULL
Adresse	smallint(6)	YES	NULL
copies	varchar(100)	YES	NULL
exemplar	varchar(100)	YES	NULL
Anmerkung	varchar(1000)	YES	NULL
Bemerkung	varchar(2000)	YES	NULL

Zer	tinyint(1)	NO	0
StatusImage	tinyint(4)	YES	NULL
StatusTrans	tinyint(4)	YES	NULL
StatusIndex	tinyint(4)	YES	NULL
OlimPrae	tinyint(4)	YES	NULL
OlimZahl	smallint(6)	YES	NULL
OlimPost	tinyint(4)	YES	NULL
viewer	tinyint(1)	NO	0
Img1	varchar(800)	YES	NULL
Img2	varchar(800)	YES	NULL
Img3	varchar(800)	YES	NULL
catalog	varchar(800)	YES	NULL
catalog2	varchar(800)	YES	NULL
catalog3	varchar(800)	YES	NULL
Edition	text	YES	NULL
Bibliographie	text	YES	NULL
LDAB	int(5)	YES	NULL
LDABurl	varchar(200)	YES	NULL

3.1.2.3. Besitzer table

The next table intends to capture a list of all holding institutions and is named the Besitzer table. It suffers from much of the same column naming complaints as above, but is otherwise sufficiently suited to its purpose:

Besitzer Table

Field	Type	Null	Default
-------	------	------	---------

BesitzerID	bigint(20)	NO	NULL
Status	tinyint(1)	NO	0
Country	varchar(1000)	NO	NULL
Place	varchar(1000)	NO	NULL
Institution	varchar(1000)	NO	NULL
Institut	varchar(255)	NO	NULL
InstitutionCurrentName	varchar(1000)	NO	NULL
Address	varchar(1000)	NO	NULL
addr_strasse	varchar(300)	NO	NULL
addr_plz	mediumint(20)	NO	NULL
Contact (person)	varchar(1000)	NO	NULL
Con_vorname	varchar(100)	NO	NULL
Con_nachname	varchar(100)	NO	NULL
Con_titel	varchar(30)	NO	NULL
Phone	varchar(65)	NO	NULL
Fax	varchar(65)	NO	NULL
Email	varchar(1000)	NO	NULL
URL	varchar(1000)	NO	NULL
URL1	varchar(1000)	YES	NULL
URL2	varchar(1000)	YES	NULL
kuerzel	varchar(10)	YES	NULL
logo	varchar(50)	YES	NULL
logo2	varchar(50)	YES	NULL

3.1.2.4. The Majority Tables

The remainder of this 450+ table database definition consists of a grouping of tables-- we will call the “Majority Tables”-- all with identical table definitions, one table per manuscript. The function of these tables is blurred; however, in general they try to capture information, sometimes about a manuscript page, sometimes about a specific image scan of that page, sometimes a transcription, sometimes a set of images, and sometimes they repeat rows per page in situations where there might be biblical content on a page which crosses book boundaries. All previous complaints apply to this table template and are multiplied. Worth noticing are: the lack of column naming consistency, the repeating ‘bild’ columns, at least 4 different usages for this table (biblical content, page information, image information, transcription information), storage of computed columns (begAdr, endAdr, bB, bC, bV, eB, eC, eV – these are used to store a single contiguous span of biblical content as both an integer with begAdr and endAdr, and those same values broken down as book, chapter, verse), an attempt to enumerate all possible features about either the page, an image scan, or the text of the page, and many fields which are specific for the VMR 1.0 usage (status, rechte, view). Possibilities to break data integrity abound and no hope is possible for future engineers to understand this table definition without reverse engineering the meaning from the data itself.

The current design imposes functional problems for the implementation of the VMR 1.0. As mentioned at the start of this [section](#), these multiple-personality

(pages/images/transcriptions/biblical content) tables are also defined with a focus on division by content, duplicated one for each manuscript. This prohibits the VMR 1.0 from performing any queries against the entire dataset without constructing an SQL statement which names individually 450 (as of now, and the code would need to change every time a new manuscript was digitized and added to the database) tables in an SQL UNION structure. The HSSListe table has 5500+ manuscripts catalogued. Imagine the SQL statement to search a completed database. Basic useful queries like, “what manuscripts contain ‘John’,” or “what manuscripts have images but no transcriptions,” are not practical. In short, any question you might asks of the data in one of these tables can only return information for a single manuscript.

A complete exemplary definition for the Majority Tables is given below with the definition for manuscript Gregory-Aland: P1:

Cont10001 Table

Field	Type	Null	Default
inID	int(11)	NO	NULL
status	tinyint(1)	YES	NULL
msNr	mediumint(5)	YES	NULL
InstID	tinyint(4)	YES	NULL
lfNr	int(6)	YES	NULL
Erg	tinyint(2)	YES	NULL
cuNr	int(10)	YES	NULL
reNr	tinyint(3)	YES	NULL
view	varchar(4)	YES	NULL

folio	varchar(6)	YES	NULL
page	smallint(6)	YES	NULL
begAdr	int(6)	YES	NULL
endAdr	int(6)	YES	NULL
bB	varchar(2)	YES	NULL
bC	varchar(2)	YES	NULL
bV	varchar(2)	YES	NULL
eB	varchar(2)	YES	NULL
eC	varchar(2)	YES	NULL
eV	varchar(2)	YES	NULL
nB	varchar(2)	YES	NULL
noNT	text	YES	NULL
ill	varchar(3)	YES	NULL
dek	varchar(3)	YES	NULL
caT	varchar(3)	YES	NULL
kep	varchar(3)	YES	NULL
hyp	varchar(3)	YES	NULL
col	varchar(3)	YES	NULL
inL	varchar(3)	YES	NULL
bIP	varchar(3)	YES	NULL
oMV	varchar(3)	YES	NULL
sSi	varchar(3)	YES	NULL
bib	varchar(3)	YES	NULL
miT	varchar(3)	YES	NULL
nTe	varchar(3)	YES	NULL
bnd	varchar(3)	YES	NULL

sca	varchar(3)	YES	NULL
bSc	varchar(3)	YES	NULL
oAV	varchar(3)	YES	NULL
inst	varchar(6)	YES	NULL
abbr	varchar(4)	YES	NULL
thumb	varchar(200)	YES	NULL
bild	varchar(200)	YES	NULL
bild2	varchar(200)	YES	NULL
bild3	varchar(200)	YES	NULL
bild4	varchar(200)	YES	NULL
rechte	tinyint(1)	NO	NULL
trans	varchar(200)	YES	NULL
bem	text	YES	NULL
date	timestamp	NO	CURRENT_TIMESTAMP
agents	varchar(1000)	NO	NULL
adr	smallint(4)	YES	NULL
test	int(11)	YES	8

In addition to the aforementioned difficulties with the table definitions, the data itself lacks systematic regularization.¹³³ Although storing this information is valuable, searching and other computer assisted research on these fields is impossible until the data is regularized. A sample of the problems follow:

Cont_nnnnn Table Excerpt

ObjID	Jh	Blattzahl	Blattzahl_Int	Zeilen	Höhe	Breite	Format
10096	VI/VII	1 Frag	1	27r,11v	27,5	16,4	ca. 16,5x9
10122	IV/V (?)	Frag	0	11	6,5	2,8	5,2x3,3 (?)

¹³³ The practice of data regularization involves formatting all of the same kinds of values in the same way, thus enabling a systematic evaluation of the data without human intuition to understand the value of the data. For example, instead of Format: ca. 16,5x9; better: heightCM: 16.5; widthCM: 9. This would allow the software system to, for example, return all manuscripts with a width between 15cm and 17cm. This is not possible with “ca. 16,5x9” as a data value. It needs human intuition.

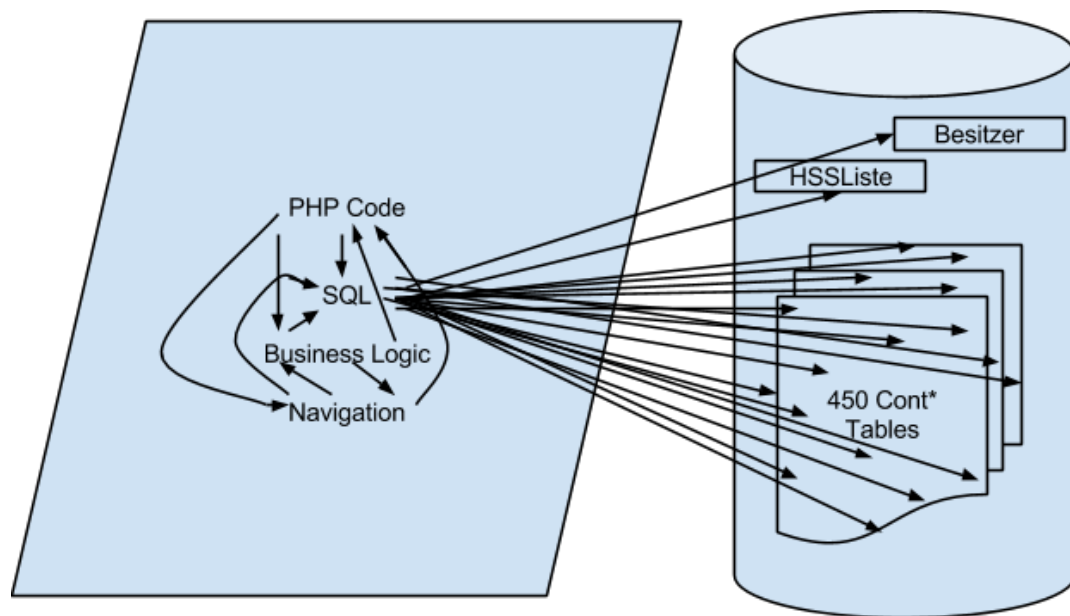
There is an initial attempt to make Blattzahl (pages) into a proper regularized column as Blattzahl Int, but inconsistencies remain in the regularization. Zeilen (line count) incorporates two values in one column for a manuscript. The Höhe and Breite (Height and Width) seem at first glance to be fine, though localized as German, but further investigation shows they obtained their values from the Format column and it is not obvious why these values are sometimes different. The larger problem is that these fields are not simply real numbers; with reference to the table definition, one will discover they are VARCHAR types, not allowing any meaningful quantity queries on these fields (without assuming a data type that differs from the data definition and then converting with a function before performing the query).

In conclusion, no computer science data modeling principles have been followed with the creation of the VMR 1.0 database, resulting in a dataset replete with diamonds buried deep in the rough of a poor data model.

3.1.3. System structure

The entire Virtual Manuscript Room functionality is implemented in PHP including all the SQL used to access the datastore. There is no tiered separation (see following section, [4-Tier architecture](#)) of business logic, from frontend logic, from the SQL necessary to access the datastore. This resolves simply to a set of PHP files which are difficult to understand by anyone except the original author of the code, which makes

the system nearly impossible to maintain moving forward. The lack of a structured design prevents other projects from utilizing any business logic embedded inside the PHP files. There is nothing which constitutes a foundation of modular component design for use as a solid base to build upon, moving forward. In short, the structure of VMR 1.0 is separated only necessarily between PHP code and the database and can be summarily represented as:



VMR 1.0

3.1.4. What happened?

Here we deviate for a moment on general thoughts regarding DH and software engineering in the academy. Today's society, and especially the commercial industries, often have an underdeveloped appreciation for academics and this seems to breed a general defensiveness, seen prominently in the humanities, which fights to establish that a professional academic is, himself or herself, genuinely qualified for their work

and provide a valuable service to the society that generally pays taxes for their work to continue.¹³⁴ As observed by the present author, whose experience of humanities research is relatively recent, this defensiveness seems to sometimes manifest as the refusal to accept help and contributions from more professionally, yet less academically experienced members, denying implicitly that there is even such a thing as expert level computer science, in the same vein as expert level Faust or linguistic studies. This can quickly devolve into a lack of respect from both sides and tarnish a working relationship between a humanities scholar and a computer scientist sharing in work together on a project. Care must be taken by the computer scientist to be inspired by the time and knowledge invested to become a leading scholar in any field, even if that knowledge does not help build a functional website. The scholar is the focus of the project. They are the end user-- the customer-- and the entire point of the software should be to make the customer happy. On the other side, sometimes the scholar needs to understand that when they offer suggestions for how software should be built, while extremely valuable for any content conveying their desired user experience, their suggestions sometimes come across a little like how one would suspect a Plutarch expert might feel if the computer scientist tried to be taken seriously commenting on Lives. The humanities scholar must ascribe the same value and respect to the expert computer scientist as they do to those scholars they cherish in their own field. All comments from both sides need to be heard with interest and respect from either side

¹³⁴ Ruth Barcan, "Why Do Some Academics Feel like Frauds?," *Times Higher Education (THE)*, January 9, 2004, <https://www.timeshighereducation.com/features/why-do-some-academics-feel-like-frauds/2010238.article>.

about either discipline-- an untarnished perspective can often yield a new idea-- even with Lives.

Beyond the issue of personal respect, a more difficult problem to solve is that many digital projects in academia, and here seen within the humanities, begin without the assistance of trained and experienced software specialists. There is often a lack of understanding and thus appreciation for the benefit such professionals can bring to a project. The concepts of longevity and maintainability elude the focus of the academic who primarily seeks for their research the antagonist of these desired traits: uncharted territories. While this mindset to experiment and explore new concepts is necessary for academic research in general, including the field of academic computer science, it proves to be detrimental to building a stable software system, to support such a research project, which has a chance to endure for decades. There is a facet of computer science training in academia which is absent from the humanities-- training for the commercial industry. True enough, exploratory research thrives in academic computer science, but a large part of training provided by the university in this field also involves preparation for professional apprenticeship within the commercial world, where computer science is seen as a means to successfully build **efficient, stable, maintainable, and secure** systems. If asked, a humanities research expert would certainly confirm the desire for all of these traits, though any expertise from computer science to provide them is often found absent from their teams. This eventually results in software systems so replete with hidden knowledge, unreadable code, and poorly designed data models that the

next humanities research expert with some enabling knowledge of software development who replaces the original author of the system eventually replaces the software system itself, out of necessity. Time spent simply reinventing the same features in new incarnations of new technologies, or simply to provide for new use cases, can be avoided with the skillset taught in academic computer science and honed in the commercial software industry, if the humanities will recognize the contribution a classically trained computer scientist can make to their digital project. Another obstacle to providing quality software development for academic projects is the academy's challenge to retain quality computer scientists due to their disparaging remuneration offered for the roles, compared to the commercial industry. It is common for a commercial software engineer to command more than two and a half times that listed on academic pay scales. According to Schmid, "In Germany it took a while before software developers were actually payed as well as researchers."

So where does that leave the typical humanities project in need of digital tools? A positive development has seen the past few year produce new "centers for digital humanities." These departments at academic institutions often have capacity to fill many of the specialist roles described earlier in [section 1.6](#)

3.2. A way forward

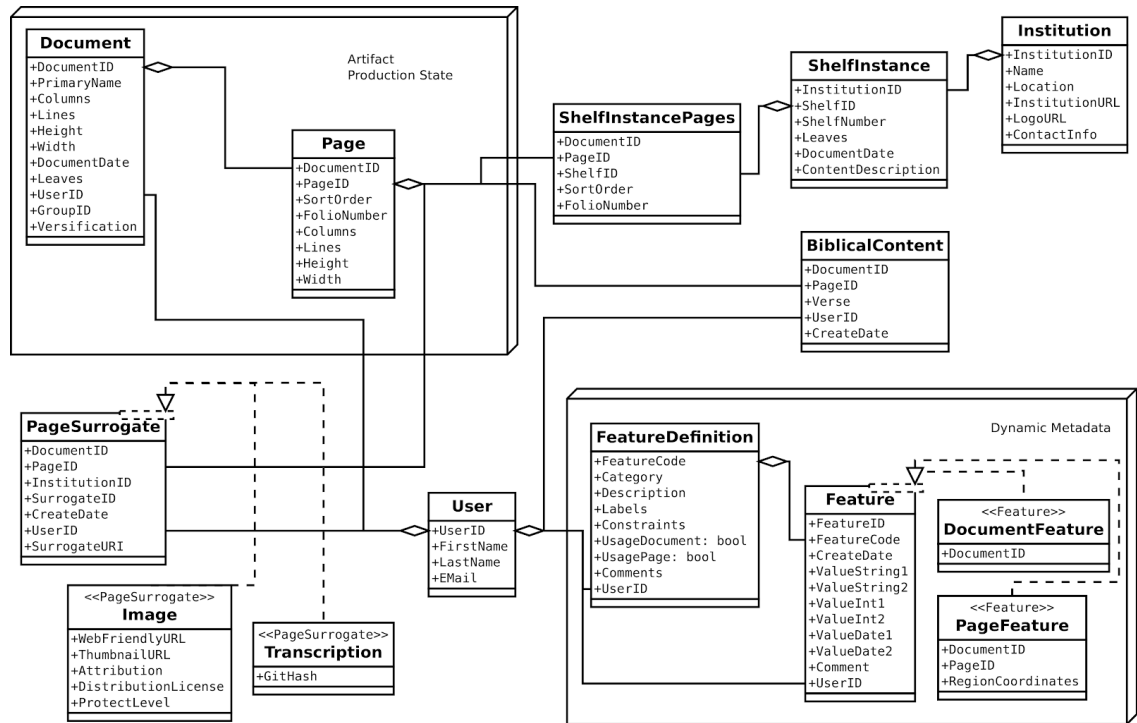
Despite all the criticisms, the initial version of the Virtual Manuscript Room succeeded in making resources available online which could previously only be examined by traveling to the INTF or to other institutions. Much necessary work was spent digitally

capturing the manuscript catalog from the printed *Kurzgefaßte Liste* and this time-consuming endeavor yielded an important datasource, even if not yet regularized and in an ideal data model. The desired use cases which can be derived from the VMR 1.0, this digital dataset of manuscript metadata, plus a limited set of both digitized images and transcriptions, are a valuable asset which should not be discounted.

To ensure these assets generated as part of the VMR 1.0 effort perpetuate forward to future generations and that upward progress is made in this effort, we seek now to lay a concrete foundation of classical computer science engineering design.

3.2.1. A normalized data model

Ulrich Schmid proved to be an ideal and willing specimen to fill the role of **business analyst** for our design efforts. Schmid is a respected domain expert who has the full faith and confidence of the Institution as a representative of their needs. Interviewing Schmid over 3 days allowed us to yield a clear and concise data model which represents exactly and only what the Institute wishes to capture in their repository. The data model follows.



The VMR CRE data model

3.2.2. The benefits of a component architecture in humanities

Workflows in humanities research are complex, yet often similar among institutes. A component architecture design promises that if these workflow complexities can be distilled into small discrete parts, building software solutions to handle these isolated tasks will produce tools better suited for potential reuse across many institutes which have similar needs. Conversely, from the perspective opposite the tool builder, if a scholar sees their workflow as one complex task, finding software already in existence to meet that task is unlikely. There are certainly obstacles to this theory: different programming languages and technology stacks in use between institutes or, if the components are distributed services in nature, where each discrete component for a

single workflow may live at different institutes-- a single system using many underlying remote services-- maintaining stability comes more heavily into play. In this case study, we will see how four universities contribute components to solve similar but different use cases required by all parties involved.

3.3. 4-Tier architecture

For the past 20 years, software engineering has embraced an n-tier-- predominantly a 3-tier¹³⁵-- application model, which delineates data storage, from business logic, from presentation. The 3-tier model was pioneered by John J. Donovan in his work at Open Environment Corporation starting in 1992.¹³⁶ N-Tier development promotes clear division of technologies and engineering skillsets. A database specialist can develop and optimize a database scheme. Software engineers can encapsulate business logic into a confined set of programming objects. Graphic designers and user interface specialists can focus on the end user experience. Software development in discrete tiers also enables better comprehensive software testing. Separate test suites can be written and easily extended as they focus to test independently the database model, the business logic, and the user interface. Bugs can be narrowed down more easily. Regression tests can assure stability as features are added to the system. Optimizations can be performed in all 3 tiers. There are many advantages to the 3-tier model of

¹³⁵ Techopedia, "Three-Tier Architecture," Techopedia, 2017, <https://www.techopedia.com/definition/24649/three-tier-architecture>.

¹³⁶ John Donovan, "Three-Tiered Client-Server Systems Gain as Enterprise-Wide Solution," *Investor's Business Daily*, October 10, 1994, https://professordonovan.com/images/pdf/three_tiered_client_server_systems.pdf.

development, and this idea has expanded in the online web development sphere over the past few years to incorporate an additional tier between the user interface and the business logic, the web services application programming interface (API). In the age of the Internet and disparate systems, a web services API tier opens up the components within a software system for use by other remote systems connected via the global Internet. These are sometimes additional instances of the same software running a partition of a project in another location, or sometimes completely different projects and software systems which find useful a specialized component of the system.

After drafting the data model for the VMR CRE, the next 5 weeks were spent framing a traditional 3-tier design plus a 4th tier web services API for programmatic exposure of the system's data and functionality to outside institutions.

During this time Schmid set out to regularize the Institute's data. Conversion scripts were created to facilitate and capture the process of migration from the VMR 1.0 datastore to the new system and to provide reproducibility for this conversion process while new data acquisition work at the INTF continued for the VMR 1.0. All software development artifacts, including design, database scripts, conversion scripts, and framework development have started their life and continue to progress within a versioned source control repository to provide traceability for the design, the migration, and new system software itself. In a very short period of time the Virtual Manuscript Room 1.0 was transformed into an extendable, maintainable, understandable system for

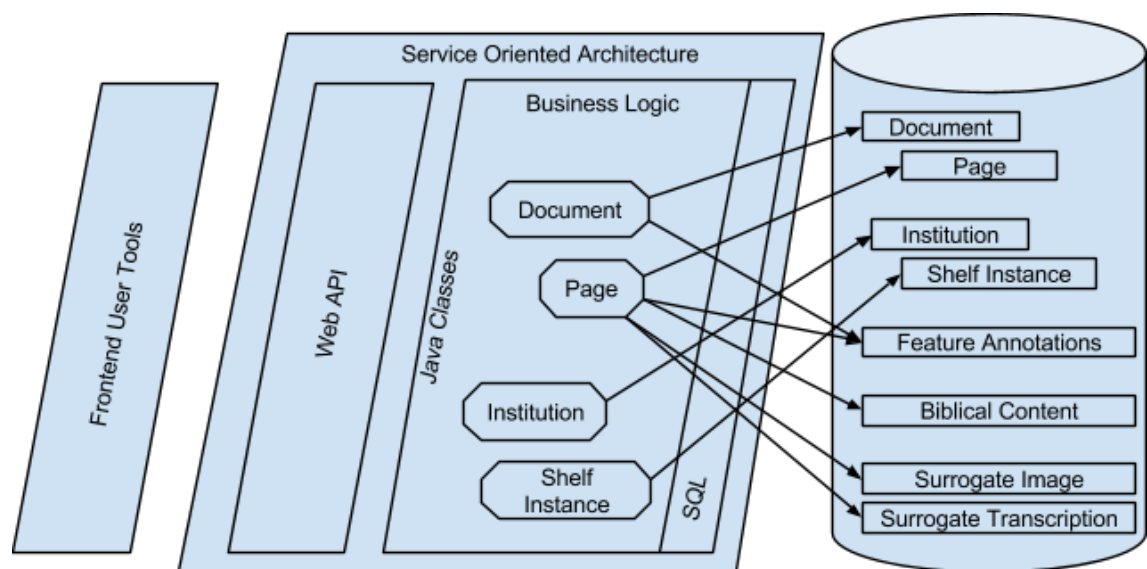
other software engineers to participate in its development and move the system forward for the foreseeable future. With proper database normalization and regularization, we have added, with no other work, the ability to generally search manuscripts and pages of every manuscript within the system based upon any combination of the original exclusive VMR 1.0 search options, plus the addition of ranges within a criteria, plus a dynamically expanding multi-faceted search fueled by our EAV design for manuscript and page feature metadata, growing as scholars add new feature definitions to the system as they desire. Some useful search facets immediately available to New Testament manuscript scholars included:

- biblical content
- line count
- column count
- page width
- page height
- document id (moved from a lookup mechanism to a true search criteria as it can be coupled with other parameters such as a biblical content range to search for specific pages limited to a specific manuscript)

As a proof of concept, to test the decoupled n-tier architecture, a mobile phone application as a second user interface (tier 4) was developed against the same web services API (tier 3) used by the desktop browser user interface. The mobile application development effort offered opportunity to better generalize the web services

tier and to a lesser extent, also the business logic (tier 2) and confirmed the intents and purposes for the full technology stack. The proof of concept shows that with very minimal code, one can reuse all the functionality built into the lower layers and expose them for new applications.

The new design, showing the clear separation of the user interface from the web services from the business logic from the datastore can be seen visually represented as:



VMR CRE

While the VMR 1.0 and thus this initial design covers only the raw materials used in the research and production of a digital edition: manuscript cataloging and metadata with images and transcriptions, these services are a vital base to build upon and useful components for any digital edition. This work molded the valuable assets produced during the VMR 1.0 effort into an extensible and sustainable platform which could act

as the foundation for future research tools produced for both the INTF and partner institutions.

Specifically, the VMR CRE has these 4 distinct technology layers:

- Tier 4, User Interface: HTML widgets
- Tier 3: Web services API
- Tier 2: Business logic: Java objects
- Tier 1: Datastore: Relational database (MySQL or similar) for manuscript metadata; file system for images; Git for transcriptions and other versioned resources.

The HTML widgets, described in more detail in [section 3.7](#), are deployed as OpenSocial gadgets within the popular Liferay web portal framework.

The VMR CRE user interface communicates with rest of the VMR CRE software system via the web services API to prove the API is sufficiently complete for other projects to utilize the entire system remotely via the exposed web services.¹³⁷

¹³⁷ The Web services API is described in detail in [Section 5](#).

Java has been used to implement the server-side system domain logic. No heavyweight frameworks were used, but instead a simple POJO (Plain Old Java Objects)¹³⁸ set of classes constitute the business logic for the VMR CRE.

3.4. Technology selection

A guiding principle has been to choose established methodologies and technologies ubiquitous in the software industry to promote the longevity of the software system.

3.4.1. UTF-8 character encoding

Traditionally, all transcription work done on the *Editio Critica Maior* (ECM) project, from both ITSEE and the INTF, was encoded for the 8-bit SPIonic font.¹³⁹ This was a reasonable choice at the time, as work on the ECM was begun long before the Unicode standard¹⁴⁰ was created. Before Unicode was adopted, the concept of a ‘character’ on a computer system was traditionally restricted to 8-bits or less. This provided, at most, ordinals for 256 glyphs, minus control codes. To allow for a greater variation of glyphs, specific fonts were created which reassigned the ordinals to the glyphs for a different language. A document would then need to specify when a font change was necessary and coordination between document content and display font was crucial to achieving proper display. Many encoding schemes were in common use to represent

¹³⁸ Naresh Joshi, “Plain Old Java Object (POJO) Explained,” Programming Mitra, 2016, <https://programmingmitra.blogspot.com/2016/05/plain-old-java-object-pojo-explained.html>.

¹³⁹ Jimmy Adair, “Keyboard Map of SPIonic,” SBL, 1998, <https://www.sbl-site.org/Fonts/readme/spionic.txt>.

¹⁴⁰ Unicode Consortium, *Unicode Standard: Worldwide Character Encoding, Version 1.0* (Reading, MA: Addison-Wesley, 1992). For the latest specification see: <http://unicode.org>

ancient Greek with ordinal values available within 8 bits (0-255). A few fonts which were created for display took hold as standard use and the ordinal mappings implemented for each of these fonts held ground as common encodings for a possible way to represent ancient Greek in an electronic document. When a document wished to change to one of the other available fonts which used a different ordinal scheme, a conversion of all Greek characters in the document would need to map from the current encoding to the encoding used by the new font. Online message communication with Greek was challenged by the diversity of encoding options and would often settle on a Latin transliteration, as did the popular B-Greek (Biblical Greek) mailing list,¹⁴¹ which began in the 1990's and continues today as the Biblical Greek Forum, now using Unicode.¹⁴² A large corpora of ancient Greek literature, the Thesaurus Linguae Graecae (TLG), adopted an encoding system developed by David W. Packard of Hewlett-Packard fame, called Beta Code.¹⁴³ Many Greek characters required more than one ordinal value in the Beta Code system. For example, an uppercase Greek character was encoded by prefixing its lowercase counterpart with the ASCII ordinal value for an '*' (asterisk); medial and final sigma were disambiguated with the ASCII ordinal values for 'S' and '1', against 'S' and '2'. Since for display this required interpretive logic to map a sequence of Beta Code bytes to font glyphs, i.e., not a one to one mapping, a simple font-only solution to render Beta Code was not possible. More than

¹⁴¹ B-Greek, "Citation of Greek Text in Messages sent to B-Greek," B-Greek: The Biblical Greek Forum, 2017, <http://www.ibiblio.org/bgreek/bgtransliteration.html>.

¹⁴² B-Greek, "B-Greek: The Biblical Greek Forum," ibiblio, 2017, <http://www.ibiblio.org/bgreek/forum/>

¹⁴³ Maria Pantelia, "The TLG Beta Code Manual," TLG, 2016, <http://www.tlg.uci.edu/encoding/BCM.pdf>.

25 years after the first published Unicode standard,¹⁴⁴ the TLG continues to use Beta Code (see the TLG Beta Code manual cited above). Some of the widely used pre-Unicode Greek font/encodings include: GreekKeys, SGreek, and SPIonic. SPIonic established a widely accepted special mapping of ordinals to Greek character glyphs which generally followed a phonetic mapping from the Latin letters in the standard ASCII¹⁴⁵ character encoding, while trying to remain as close as possible to the Beta Code encoding system. As mentioned initially, the SPIonic font mapping was chosen for INTF and ITSEE transcription projects before the popularization of Unicode.

The Unicode initiative sought to assign a unique ordinal value to every glyph of every language in the world. This meant that no longer did a specific font need to be tied with every character in a document. It also meant that each character within a document could no longer simply be stored within 8-bits, since ordinal values above 255 were assigned by the Unicode Consortium. Simply allocating 4-bytes (32-bits or approximately 4.2 billion ordinal possibilities) for every character – the minimum number of bytes required to represent any ordinal value assigned by the Unicode standard – would be grossly wasteful, nearly quadrupling the size of all existing documents. New ways of encoding a ‘character’ were developed, including the now dominant UTF-8 encoding. UTF-8 reserves the first bit of every byte in a document to

¹⁴⁴ Unicode Consortium, *Unicode Standard: Worldwide Character Encoding, Version 1.0*. (Reading, MA: Addison-Wesley, 1992).

¹⁴⁵ American Standards Association, Inc., *American Standard Code for Information Interchange*, New York NY, June 17, 1963; <https://web.archive.org/web/20170701043712/http://www.worldpowersystems.com/J/codes/X3.4-1963/>.

designate if the byte represents an entire character (0), or if it is part of a multi-byte character sequence (1). ASCII was given privilege in the Unicode specification and retained its entire ordinal mapping; ordinal values 0–127 have the same representation in both ASCII and Unicode. Since these values only require 7-bits to represent, always leaving the initial bit in every byte as a 0, UTF-8 becomes backward compatible with ASCII documents. This means all ASCII-only documents can remain unchanged and be considered UTF-8 encoded Unicode. If more than 7 bits are needed to represent the ordinal value of a character, multiple bytes are used, with the first bit of all bytes in the byte sequence raised, signaling every byte as part of a multi-byte sequence. The second bit of each byte in the sequence then differentiates the first byte in the sequence (1) from the subsequent bytes (0). The first byte in the sequence specifies the count of total bytes in the sequence by consecutively raising one more bit for each byte beyond two. For example, if a character's ordinal value requires 2 bytes of storage to represent, then the initial byte will begin with the 3 bits: 110. The first bit designates this byte is part of a multi-byte character sequence. The second bit designates that this is the very first byte of a multi-byte character sequence, and the count of raised initial bits ('11') means that the entire multi-byte character sequence consists of 2 bytes. If a character's ordinal value requires 3 bytes of storage, then its initial byte would begin with the 4 bits: 1110. The remaining bits of this first byte and all bits, save the first two to designate participation in a multi-byte character sequence, of the subsequent bytes of this character sequence, represent the actual character ordinal value. This means that only the number of bytes required to represent a character are used and consequently less

bytes are required to represent characters from languages assigned to lower ordinal numbers in the Unicode specification. Greek requires a maximum of 2 bytes for any character. Today UTF-8 is the dominant encoding for Unicode data on the Internet.¹⁴⁶

A conversion tool was developed by this project which allowed easy migration of existing transcriptions from the legacy SPIonic encoding to UTF-8 encoded Unicode. This utility was written in Java and runs on macOS, Linux, and Windows. Since macOS is the dominant platform at the INTF and ITSEE, an AppleScript wrapper around the tool was developed to allow a user to simply drop an SPIonic encoded text file onto the utility's desktop icon and the file will be converted from SPIonic to UTF-8.¹⁴⁷

3.4.2. TEI

The Text Encoding Initiative began in 1987, with work organized initially by the Association for Computational Linguistics, the Association for Computers in the Humanities, and the Association for Literary and Linguistic Computing, to overcome the barriers presented by the then diverse and often proprietary data formats used across projects in the humanities. Today, one will be hard-pressed to find a project in the humanities which does not, in some way, store or offer their data using the TEI. In this respect, the TEI has been a success. The Extensible Markup Language (XML), in

¹⁴⁶ Q-Success, "Usage Statistics of Character Encodings for Websites," W3Techs, 2014, http://w3techs.com/technologies/overview/character_encoding/all.

¹⁴⁷ The source for the conversion utility can be found here: <http://crosswire.org/svn/community/trunk/utls/java/TranscriptionReconciliation/src/org/crosswire/mss/trascript> and the compiled macOS drop target application here: <http://crosswire.org/ftpmirror/pub/vmrcr/utls/sg2u.app.zip>

which TEI is defined, has this same goal and selecting XML went a long way to ensuring longevity and data exchange between institutions by ensuring data was stored in a human-readable format. Further to these, TEI also hoped to encourage common tool development across a global set of data all using a common markup. Most editors of the TEI specification and the projects which use TEI agree that TEI has not achieved this anticipated level of interoperability.¹⁴⁸ The TEI Consortium defines a superset of all tags conceivable for text encoding and recommend that a project select a subset of these tags which cover the needs for the intended project task. This has led to projects selecting a wide variety of subsets. The element usage is also defined with a degree of freedom to gain better buy-in from humanities projects, which has been effective; as stated initially, one would be hard pressed to find a digital humanities projects not using the TEI. This freedom, however, has caused variation in the way in which each tag is applied across projects, contributing to the lack of interoperability now widespread. There are two additional major contributing factors to this reality, as well. First, there is the absence of ‘that killer app’. There is a void of major software products which are essential for a humanities scholar to do their research which use the TEI specification. This is not to say that tools do not exist which operate on TEI; most digital humanities projects have built at least one. Or that there are no major editors which directly support editing TEI markup; Oxygen and others do this excellently. What is missing are a plethora of tools which are indispensable to the (non-digital) humanities scholar to do their research (tools which incidentally use the TEI standard). The presence of a

¹⁴⁸ Desmond Schmidt, “Towards an Interoperable Digital Scholarly Edition,” *Journal of the Text Encoding Initiative*, no. Issue 7 (November 12, 2014), doi:10.4000/jtei.979.

dominant application providing substantial benefit to the scholarly community, and which works behind the curtain against its interpretation of the TEI, would cause projects which utilize this beneficial application to naturally conform their usage of the TEI to accommodate the application to achieve their desired result from the product. As an example, this can already be seen from the suite of tools produced for the Perseus project. Collaborating institutions which desire their data to be included in the Perseus Digital Library must conform their use of the TEI standard to the strictly defined guidelines prescribed by Perseus-- and with their software. If an institution finds that their text does not display properly in Perseus, they will adjust their markup to get the desired result. The software is a catalyst for conformity.

3.4.2.1. WYSIWYM

Semantic markup, like the TEI, dominant in humanities, shifts focus from **display** to **meaning**; e.g., more importance is placed on marking a missing segment of text in a transcription of a manuscript as "no longer present; damage to canvas material; text supplied by transcriber from Nestle-Aland, 28 ed.," over the visual display of that segment of text in the editor in [brackets]. What-you-see-is-what-you-mean (WYSIWYM) text editors attempt to easily allow authors to attach semantic meaning to segments of the text they edit. When semantic meaning is ascribed to a portion of text, these editors show visually where text is annotated with meaning. The visual display in which WYSIWYM editors choose to display semantic meaning is not necessarily the final display of a published transcription; it is simply some useful visual demarcation, often familiar to an editor, of all meaning tagged within the transcription.

A second factor which has contributed to the lack of unity in TEI markup usage is the absence of quality, free, what-you-see-is-what-you-mean (WYSIWYM)¹⁴⁹ text editors which generate TEI for the scholar. Humanities scholars should not be required to learn and use pointy-bracket markup. The practice of training scholars to hand edit TEI markup as they transcribe a manuscript has, in large part, directly caused the TEI interoperability we see today, to the extent that even individual members of the same project use markup semantics differently. A quality, specialized editor which provides the full functionality necessary for a project's domain, and which automates the TEI markup on the backend would cause consistent markup among all participants in the project. Further, if multiple project chose to use the same editor, TEI usage among projects would be much more consistent. This might be conducive to the development of 'that killer app', which could work across projects, against a much larger uniform dataset. Finally, if both an editor and a killer publishing/research tool agreed interpreting the TEI such that they could be used in tandem, usage of the TEI would begin to unify across the humanities in academia.

WYSIWYM poses a unique problem for humanities scholars. TEI defines a **semantic** markup for elements of a text; display directives are discouraged. The scholar should see in the WYSIWYM editor that what they edit has the semantic meaning they intend. The way the WYSIWYM editor displays this meaning is not intended to represent how

¹⁴⁹ "What you see is what you mean" vs. the more common, "what you see is what you get."

a text will or should be displayed in the end to a reader. This needs to be made clear to the transcriber; it is the observation of this researcher that text scholars seem exceptionally concerned with how their text will be displayed. The WYSIWYM editor's display need not match how the transcription is displayed to a reader.

3.4.3. Git revision control system

Revision control systems have been used to track changes on computer systems and for software development projects since the early 1980's. A revision control system allows multiple versions of the same file to be saved, along with notes about each revision, when a revision was made and by whom, and can show exact changes between versions. A concept known as branching also allows more than one history of a file to be maintained, for example, a 'stable 1.0' branch of the file which only receives bug fixes (1.0.1, 1.0.2, etc.), and a 'development' branch of a file where new features are being added. The first revision control systems, such as RCS, could manipulate a file only by a single user at a time. Difficulties ensued when multiple development team members would work on the same file simultaneously. Both users would begin editing a file at revision n . The first user would save changes, creating a new revision $n+1$, but when the second user saved changes and created $n+2$, the changes in $n+1$ would be lost; the desire would be for both changes to be included in revision $n+2$. The initial solution to this problem was to build locking mechanisms into the revision control system, allowing one user to check out and lock a file while working, and thus not allowing any other developer to make changes during this time until the file was saved

and the lock released. While this avoided the problem scenario, it did not ultimately accommodate the desire for multiple users to simultaneously work on the same file, thus a new breed of revision control systems was born, the concurrent revision control system, pioneered by tools like the Concurrent Versions System (CVS)¹⁵⁰ in 1990 and later Subversion (SVN)¹⁵¹ in 2000. These tools introduced a file patching mechanism along with a client / server architecture which kept an authoritative copy of the repository on a server, with multiple developers checking out their own copy of the repository at the latest version, without the need to lock any files. Change could be made by all developer to their local copies of files and when editing was finished, all the changes would be committed back to the server. In our problem scenario, two users check out the repository at revision n , both users change the same file. The first user checks in changes, incrementing the repository to revision $n+1$. The second user finishes editing and attempts to commit changes. At this point the concurrent revision control system informs the user that a commit cannot yet be performed because the user does not have the latest version of the repository. This user must perform an ‘update’ operation which is the newly added feature to a concurrent revision control system designed to solve the concurrency issue. An update will retrieve all differences between the last version of the repository which a user has checked out (in our case, our second user has version n) and the latest version of the repository (version $n+1$) and apply those changes to the user’s local copy of the files. In our scenario, these are exactly the changes made by our first user. The changes are applied line by line

¹⁵⁰ Free Software Foundation, “CVS - Concurrent Versions System,” CVS, 2017, <http://cvs.nongnu.org>.

¹⁵¹ Apache Software Foundation, “Apache Subversion,” SVN, 2017, <https://subversion.apache.org>.

automatically by the system and as long as the exact same line in the file was not changed by both users, an automatic update can be performed to bring our second user's local copy of the repository up to date. In the unfortunate case when any of the same lines were indeed modified by both users, then those lines would be marked and the file would be put in a 'conflict' state forcing our second user to examine the lines in question and manually merge her own changes with those changes coming from the update. Once the update succeeds with no conflicts remaining, the second user's local copy of the repository is now baselined at revision $n+1$ and a commit would be allowed, creating revision $n+2$ with both changes now present in the latest version of the file. Concurrent revision systems dominated the software industry for 15 years, until the next wave, the distributed version control systems (DVCS) gained popularity. Today, Git¹⁵² is the dominant DVCS system and has the advantage that no central server is necessary to maintain a repository. Instead of checking out a copy of the latest version of all files in a repository from a central server, as done in a concurrent revision control system, a user instead clones the entire repository-- every revision of every file-- from any other clone of the repository and can work locally, creating revisions and branches, all committing to their own entire repository, locally on their computer. When work from two users needs to be merged together, DVCS systems provide a means to apply entire changesets between repositories using a concept called 'push' and 'pull'. Pulling is roughly similar to the update command described previously for keeping a local copy of the files up to date in a concurrent revision control system, and

¹⁵² Git Community, "Git," git-scm, 2017, <https://git-scm.com>.

pushing is roughly analogous to committing local changes to the authoritative server; however, in a DVCS, no one single authoritative server necessarily exists. Changes can be pushed and pulled directly between multiple developer. DVCS systems are sometimes also called decentralized version control systems. A popular free hosting website used for sharing entire Git repositories is called GitHub¹⁵³ and can be used by anyone wishing to publicly share versioned files with Git.

Git has been chosen as the tool to save and keep track of changes to transcription files and other versioned project data within the VMR CRE.

3.4.4. MySQL relational database management system

Relational database management systems (RDBMS) have been the workhorse for data storage and retrieval beginning with Codd in the 1970s. RDBMSes have been highly optimized and hardened over 40 years to accommodate for the most demanding and sensitive datastore needs. MySQL was originally developed as an open source RDBMS by Michael Widenius and was later bought by Sun Microsystems, which in turn was purchased by Oracle-- one of the largest manufacturers of commercial databases to date. MySQL has been chosen to house data repositories for Facebook, Flickr, Twitter, and YouTube, and has also been chosen to implement the relational data model for the VMR CRE. An XML, JSON, or other NoSQL data store was not chosen for this project. These technologies have advantages in distributed data architectures; they achieve this by essentially removing the complexity and power of the relational

¹⁵³ GitHub, Inc., “GitHub,” GitHub, 2017, <https://github.com>.

data model. Much of our data is relational, none of our data needs to be distributed across a processing farms for massively parallel processing-- which many of these databases were originally written to solve. Our relational data model is clearly defined within a relational database and where we need extendable data storage, our EAV¹⁵⁴ model serves to both clearly define and also avoid yet another technology dependency.

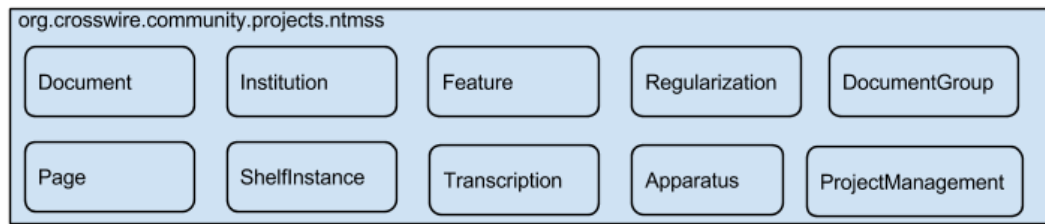
3.4.5. Java

Java¹⁵⁵ has been chosen as the server side programming language in which to solidify our business logic. According to the TIOBE programming language popularity index in August 2017, Java holds a 12.96% popularity rating at first place, before C/C++ with a combined 12.27% and then the field goes quickly downhill with at C# next at 4.19%, and Python at 3.68%. Popularity does not equate to technically superior. Those are debates left to programming language theorist and every other software developer sitting home at a Friday night LAN party. But what popularity does bring is a better chance this system will be supported long into the future, with a larger pool of able developers who might contribute to its improvement.

¹⁵⁴ See [3.6. Tagging features](#) and [3.1.2.2.1. Entry Attribute Value](#)

¹⁵⁵ Oracle, "Java + You," Java, 2017, <https://java.com>.

A view of the VMR CRE business logic objects is shown below.



VMR CRE, Business Logic, Java Objects

3.4.6. Web services API

While not strictly a technology per se, but rather a design choice implemented using standard web browsing technologies, a web services application programming interface (API) provides programmatic access to services of a computer system. An API is designed to offer programmers a chance to develop unique applications which utilize the functionality of an existing system. A web services API makes the facilities of a software system available over standard internet HTTP protocols which are the backbone of modern Internet web browsing. The VMR CRE web services API exposes the complete functionality of the VMR CRE programmatically for use by other projects wishing to use the tools or contribute to the dataset of a project running the VMR CRE software system. Web service APIs in general not only open access to a software system, but can also be executed remotely, from an environment across the globe.

The concept of a web services API is born from a long history of technology which asks a remote computer to execute some function, and goes back to the 1980's, with

various incarnations since. Originally called a remote procedure call (RPC), some usages are as popular as the Network File System (NFS) specification¹⁵⁶ still in heavily use today, implementing commands to manage files on a remote computer systems as if they were local. With the advent of Object Oriented Programming (OOP), popularized by the C++ programming language in 1983, the Common Object Request Broker Architecture (CORBA) became a means to work with remote objects, enabling remote object instantiation and method invocation.¹⁵⁷ The end of the 1990's saw a surge in XML interest and the Simple Object Access Protocol (SOAP) was developed at Microsoft as a standard to use XML to send commands to remote machines and receive the result. SOAP was adopted as a standard by the W3C with version 1.1 published in 2000.¹⁵⁸ Java brought Remote Method Invocation (RMI),¹⁵⁹ and most recently JavaScript has brought the concept of Asynchronous JavaScript and XML (AJAX) which allows a webpage to call back to the remote server to execute specific functions and update a user's display with the results. AJAX is the beginning of modern web services APIs and is described in more detail below.

3.4.6.1. AJAX and CORS

Initially, for security reasons, a webpage was restricted by a web browser from calling remote functions on any server but the one from where the webpage itself was loaded.

¹⁵⁶ Russel Sandberg, "The Sun Network File System: Design, Implementation and Experience" (in Proceedings of the Summer 1986 USENIX Technical Conference and Exhibition, 1986).

¹⁵⁷ Object Management Group, Inc., "CORBA - Common Object Request Broker Architecture," CORBA, 2017, <http://www.corba.org>.

¹⁵⁸ W3C, "Simple Object Access Protocol (SOAP) 1.1", W3C, 2000, <http://www.w3.org/TR/2000/NOTE-SOAP-20000508>.

¹⁵⁹ Oracle, "Remote Method Invocation," Java - RMI, 2016, <http://www.oracle.com/technetwork/java/javase/tech/index-jsp-136424.html>.

For example, a webpage from <http://manuscriptroom.org> could not contain any code which would be allowed to call a remote web service at <http://crosswire.org>. While not defined by any specification, all modern web browsers follow this rule to prevent one website from writing code to maliciously reference another website. This concept, while not strictly defined as any specification is called the Same Origin Policy.¹⁶⁰

Despite the legitimate security concerns, the usefulness and tradition of remote procedure calls has caused web programmers to find workarounds to thwart these browser restrictions. The use of IFRAMES and the JSONP¹⁶¹ method allow requests to contact origins other than a webpage's original source. Only recently has the W3C officially published a recommendation which specifies how a web service running on one server might advertise that it is safe and welcomes calls from other locations. This recommendation is call Cross-Origin Resource Sharing (CORS)¹⁶² and is now implemented by all modern browsers. CORS has been an important advancement, bringing legitimacy for the concept of remote procedure calls to the World Wide Web, and thus legitimacy for web services APIs. Without CORS a web services API would not have a W3C approved means to offer functionality running on one domain for use by a web page loaded from another. In other words, CORS allows a webpage on <http://crosswire.org> to contain code which legitimately makes a request to <http://manuscriptroom.com>-- the very reason web services APIs exist.

3.4.6.2. REST

¹⁶⁰ W3C, "Same Origin Policy," W3C, 2017, https://www.w3.org/Security/wiki/Same_Origin_Policy.

¹⁶¹ W3Schools, "JSONP," W3Schools - JavaScript, https://www.w3schools.com/js/js_json_jsonp.asp.

¹⁶² Anne van Kesteren, "Cross-Origin Resource Sharing," W3C Recommendation, 2014, <http://www.w3.org/TR/cors/>.

REST is a design convention for building web services APIs which has gained traction in recent years. REST essentially attempts to standardize each web services API function as an operation of either create, read, update, or delete (CRUD) persistence. While REST is not strictly a standard, only a concept, it has gained ground in many corners of the Internet and has given a similar feel to various web services APIs which simply persist data. REST does not attempt to describe how to expose remotely an entire application programming interface not centered on object persistence, though many projects have tried to use it for such.¹⁶³

The VMR CRE exposes the full extent of its capabilities via its web services API. Its capabilities are not relegated primarily to the reading, writing, updating, and deleting of persisted objects, thus, the VMR CRE has not chosen to follow REST conventions for its web services API but has chosen naming conventions which will be familiar to developers comfortable with REST.

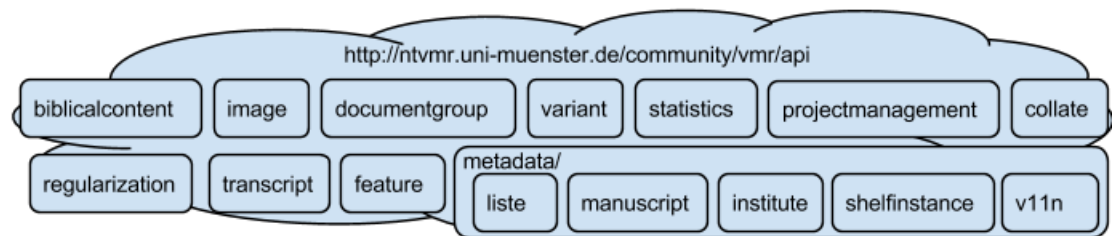
3.4.6.3. The VMR CRE web services

The VMR CRE web services API is accessible over HTTP/HTTPS and specifies full Cross-Origin Resource Sharing (CORS) access in each endpoint header, advertising to a web browser that any web page loaded from any domain may safely make requests to the VMR CRE web services. The API can be found and browsed from the following URL:

<http://ntvmr.uni-muenster.de/community/vmr/api>

¹⁶³ Troy A. Griffiths, “Why REST Sucks”, VMR CRE, 2014, <http://vmrcr.org/web/scribe/home/-/blogs/why-rest-sucks>.

Exploring the API via a web browser is supported and each API function, or “endpoint” includes a full usage page documenting its input parameters and operation. A diagram of the available services can be seen in this figure:



VMR CRE Web Services

For a detailed review of the web services API provided with the VMR CRE, see [chapter 5](#).

3.5. User interface

Building the VMR CRE as a collaborative online environment has dictated that the Web and its technologies be our primary interface to our user community. Much has already been said about isolating development efforts to components specifically for the humanities which are hard to find ubiquitously on the Internet. An overwhelming majority of most any project’s digital needs are not humanities-specific, and for these needs, finding existing software is necessary for success. This project chose the popular open source portal, Liferay, to satisfy the non-humanities needs of the project. Liferay offers management of user and of teams, shared calendars, blogs, easy content management for drag and drop, component-oriented website construction, and hundreds

of other components the teams which will use the VMR CRE system may need.

Choosing an established and comprehensive solution for the web presence of our community portal gives the future users of our system options for their project which we may not have anticipated, assures the technology choice will not inhibit the longevity of the project, and most importantly lets this development team focus on the tools specific for our domain, which other popular tools likely will not provide.

3.5.1. Use of HTML widgets

The VMR CRE user interface is built as a collection of independent HTML widgets which interact in an ecosystem using publish / subscribe messaging (see next section, [3.5.2. Inter-Gadget communication](#)). The HTML widget specification currently used is a framework developed by Google in 2008 called OpenSocial gadgets. There are multiple portal frameworks that support this HTML widget specification, including Liferay, which is our primary deployment platform. OpenSocial provides a solution for 3 primary deficiencies in other research systems:

- **Lightweight, standards-based component architecture.** Writing an OpenSocial gadget is easy. Turning an existing HTML-based tool into an OpenSocial gadget requires about 10 minutes of work; any HTML/CSS/JS blurb can be published as an HTML widget simply by adding a basic OpenSocial XML header. The entry point for a developer is low, and unintrusive to the work a developer has already accomplished-- provided their current development efforts are web-based.

- **Programming language and platform agnostic.** Every programmer has their preferred technology stack. The diversity of Java, Python, Perl, Ruby, and PHP between institutes makes sharing components difficult or even against policy if a department has standardised on a technology. The one thing all browser-based projects have in common is HTML and JavaScript, and this is where the OpenSocial HTML widget standard is defined. OpenSocial is agnostic of the server-side programming language used in deployment. For example, OpenSocial does not care or even know if your gadget was delivered from your server with PHP, Python, or Java.
- **Remote Distribution.** An OpenSocial portal simply requires a URL reference to a gadget for that gadget to be 'installed' into an online workspace (= OpenSocial container). The workspace does not care if that URL points to the same server on which the workspace runs, or to a server at a different remote institution. This helps facilitate collaborative, distributed systems by allowing a department to publish HTML widgets specific to their specialty, making their expert tools available for use externally by anyone running an OpenSocial container as their workspace.

Essentially, component reuse has already begun once an organization chooses a portal in which to deploy their HTML widgets. Typically, OpenSocial-enabled portals include hundreds of other gadgets already available for community collaboration: Message Boards, Chat, Wikis, File Sharing, Planning, Task Management, Rights and Roles,

Calendars, and much more. These are components which a humanities team will not be required to write themselves, and though not specifically tools for humanities research, these tools can greatly enhance an online collaborative research experience. This frees up a team to focus efforts on their primary objectives: humanities research tools. If a team later decides that a chosen portal environment is not ideal for their project, their humanities development work is not wasted, as they may choose another portal in which to deploy their HTML widgets.

Some components available in the VMR CRE include:¹⁶⁴

- **Document Catalog** with dynamic annotation tagging and multifaceted search.
- **Image Viewer**, including drag to pan, two-finger zoom, on-image annotations marking, and permalink URLs for publishing a desired view and annotation of a page to others.
- **Transcription Viewer**, displaying visually appealing by-page TEI transcriptions.
- **Collation and Variant Graph** components for dynamic comparison and visual representation of differences across multiple witnesses to the same text.
- **Regularization Editor**, allowing a scholar to build up their own rulesets defining which differences between witnesses matter to them and which

¹⁶⁴ Each component of the system is detailed in [chapter 4](#).

differences can be regularized out of the difference list and excluded from the display.

- **Transcription Editor** which provides easy WYSIWYM tools for non-technical scholars to richly markup a transcription with TEI attributes, without exposing them to pointy brackets.
- **Indexing Facility** for building up content metadata regarding an image library.
- **Image Management** tools which provide visual drag and drop association of images to document pages, or regex pattern matching for mass image set handling, including automated thumbnail generation.
- **Project and Volunteer Management**, allowing users to claim responsibility for, or be assigned to discrete components of work on a document: indexing, transcribing, etc., and crucial overall status views to show at-a-glance a project's progress, and what work is available for where a volunteer might begin to contribute.
- **Apparatus Editor**
- **Project Status Overview** and **New Work Contributed** feeds, showing an active, vibrant project and rewarding volunteers with visibility for submitted work.

All of these tools were necessary to implement the digital workflow at the Institut für Neutestamentliche Textforschung and are available for solving similar tasks in other projects. Many of these components can be used without a project setting up and

maintaining their own copies of the tools, as they can be included remotely from the INTF into any OpenSocial-enabled workspace.

3.5.2. Inter-Gadget communication

HTML Widgets perform their individual tasks, but a sense of disjointedness will hover over the user interface as a whole if there is no interaction between the components.

Gadgets need community too. OpenSocial's publish and subscribe (pub/sub) framework was harnessed to solve this problem. Designing software in a pub/sub model often inverts the traditional logic used for application engineering. Instead of one component instantiating another and invoking a desired method on that instance (Catalog gadget: "Hey image viewer next to me, show this image." More on this below), in a pub/sub system, components publish events which happen within themselves which they feel might be interesting to other components. They do not dictate or even know what components will listen or what they will do. Interested components subscribe to events they desire to hear about. For example, in a traditional system, a component which displays a list of folios for a manuscript might perform this action:

When a user clicks on a folio, instantiate an image viewer and invoke the display method on that image viewer, passing the URL of the image to display.

In a pub/sub model:

When a user clicks on a folio, three events are published: 'Folio Selected (docid, folioid)', 'Image Selected(url)', 'Biblical Content Selected(indexing data)'.

This component remains agnostic as to what might result from these announcements. An image viewer component could co-exist on the same page which would subscribe to the ‘Image Selected(url)’ message and would then display an image to the user whenever this event was published.

This pub/sub model enables extensible and creative ways components can be arranged to interoperate, and allows new components to be added into the mix without making existing components aware of their new presence.

Pub/Sub models are by no means new. They have traditionally been relegated to the domain of Graphical User Interfaces (GUI), with events such as: buttonClicked, mouseMoved, windowScrolled, etc. allowing programmers to perform actions based on these events. What is novel about the OpenSocial design is that components are typically equal peers of each other. Each component may both publish their own events and also subscribe to events from peers. Software developers writing code against a GUI system typically only subscribe to events and it is less common that they would publish their own.

The HTML Widgets in the VMR CRE will use this OpenSocial publish and subscribe event messaging model to give a more cohesive feel to the workspace of individual components and allow users to creatively mix and match components in a way that makes sense for their desktop. Humanities-specific events will be identified and when these events which occur within a gadget they will be **published** to the world: “the user just selected manuscript 456,” “the user just added an annotation box on an image,”...

Other humanities gadgets will listen for (= **subscribe** to) these events which they care to receive. Coordination needs to be considered when creating gadgets which are designed to work together. Many of the VMR CRE gadgets were written in view of collaboration under the EU Interedition Project banner and use a messaging namespace appropriate. For example, the VMR CRE Image Viewer gadget listens for a message:

```
interedition.image.selected(url)
```

The VMR CRE Catalogue gadget publishes this message along with the image URL whenever a user clicks on an image thumbnail. If a new gadget is written to work within the VMR CRE ecosystem, this new gadget will need to publish and subscribe to messages appropriate. For example, if a new tool would like to utilize the VMR CRE Image Viewer gadget within another portal workspace, one would need to publish the expected `interedition.image.selected` message.

3.6. Tagging features

Collaboration by an extended body of researchers is a key goal for the software platform. To encourage this, it is necessary to include a natural mechanism for scholars to add material to the system while they are performing their research. This mechanism must provide useful functionality to the contributor as an incentive to encourage participation.

The concept of 'tagging' on the global Internet is used to form a relationship between an object and a concept or category. One might apply a “family” tag individually to all photographs of one’s family uploaded to the popular photo service Flickr,¹⁶⁵ or associate a “TEI” tag to each article found online and bookmarked on Delicious¹⁶⁶ which deals with the Text Encoding Initiative. Flickr and Delicious use the tagging concept to easily associate user defined metadata enabling their users to then easily find all objects associated with their tags: images of their family or all articles relating to TEI.¹⁶⁷ Multiple tags can be associated with an object and multiple objects to the same tag. Tagging is simple, flexible, and useful for the user. The flexibility for users to choose any name for a tag can result in different names being used for the same concept. The VMR CRE incorporates the concept of tagging, with three distinctions. First, the term “feature” was more comfortable to the text researchers at the INTF, and thus is the term used for the tagging concept in the software system: a manuscript has a feature, rather than a manuscript has a tag. Next, features need to be defined before they can be used. The INTF has defined a set of features which they find useful to associate with objects in the system. Users can define additional features if they desire to tag something which they do not find in the feature set provided by the INTF. When assigning a feature tag to an object in the system, it must be selected from the list of defined features-- defined either by the INTF or personally defined by the user. This encourages the use of the same feature name for the same concept. Next, when features

¹⁶⁵ Oath, “Find your inspiration. | Flickr,” Flickr, 2017, <https://www.flickr.com>

¹⁶⁶ Maciej Ceglowski, “Delicious,” del.icio.us, <https://del.icio.us>

¹⁶⁷ Adam Mathes, “Folksonomies - Cooperative Classification and Communication Through Shared Metadata,” Computer Mediated Communication, 2004, <http://adammathes.com/academic/computer-mediated-communication/folksonomies.html>.

are defined, they are given a category. This provides organization for the feature set. The final variation on the tag concept within the software system comes with the possibility to also include one or more values when associating a feature to an object. What values can be provided, along with the labels and types of those values are declared as constraints for the feature when it is defined. For example, one might attach the feature, “Liturgical” to a manuscript object and with the feature, provide the liturgical type as “Weekdays”. This allows for a researcher to query for all “Liturgical” manuscripts or if desired more specifically, “Liturgical, type: Weekdays”. A user might wish to highlight on a manuscript page image a particularly odd καὶ compendium, tagging as a “Grapheme” feature and supplying a value of: “καὶ”. This would allow a paleography instructor to select all καὶ compendia tagged within the system as examples for their students. Feature tagging is integral for enabling scholars to collaboratively participate in contributing data. Including the concept of feature tagging within the system provides value to the contributor as incentive to add data, promoting both natural growth and pruning of the community maintained dataset, with the condition features are clearly defined before they can be used.

To implement this concept of feature tagging within the system, the Entity-Attribute-Value (EAV) construct recommended earlier (see [section 3.1.2.2.1](#)) will be utilized.

3.6.1. RDF

The Resource Description Framework (RDF) is a specification for exchanging data on the Internet. In short, RDF prescribes forming statements and relationships about a entities using triples: subject-predicate-object statements. For example subject(orange)-predicate(is a)-object(fruit). Triple browsers and datastores have been built around this simple description concept. Ontologies have been defined for various domains, providing recommended vocabularies to use with this syntax. Some might suggest RDF as an alternative to the feature tagging system we have described in this section. RDF was not chosen for a number of reasons. First and foremost, it brings no material benefit to this project. RDF certainly could have been used to store feature tag data-- even all data-- for this project, but the advantages of doing this are not evident. SQL databases are mature, fast, and most importantly, standard. RDF is a recommendation for data exchange. Technologies which store RDF are young, not as efficient in relational scenarios, not standard. RDF will be offered as a result format from the web services API if a genuine request for such format ever arises. For more on RDF and the real problems of inter-organization data exchange, see, “RDF is Meaningless”.¹⁶⁸

¹⁶⁸ Troy A. Griffiths, “RDF is Meaningless, ... or Inter-institutional Collaboration,” VMRCRE, 2016, <http://vmrcrc.org/web/scribe/home/-/blogs/rdf-is-meaningless>.

3.7. Rights and roles

Rights (the permissions to perform a discrete operation within a software system) and roles (collections of rights into groups, for assignment to a user) in the VMR are implemented using the rights and roles functionality of the Liferay portal. Liferay already has a rich and mature framework and clean user interface for establishing and assigning rights and roles. The roles in the VMR, along with their rights are as follows:

- **VMR Administrator** – This role allows full access to all parts of the system, including exclusive rights to create and edit public manuscript objects.
- **Expert User** – An expert user has access to view images which have been restricted for viewing by expert researchers only.
- **Internal User** – Like the Expert User role, an Internal User gains access to view images which are restricted for viewing by only internal organizational members.
- **Index Manager** – The Index Manager is allowed to correct index entries by other users. Without this role, a registered user will only be allowed to add new index data and to edit their own entries.
- **Transcription Manager** – A Transcription Manager gains the rights to publish a transcription for public viewing as the official transcription for a manuscript page.

3.8. Scope

The scope defined here reflects the project's initial concept and objectives. Reflecting at the conclusion of development, these core goals have served as a focal point to keep feature creep¹⁶⁹ in check. Contributing to the New Testament Virtual Manuscript Room (NTVMR)¹⁷⁰ community's early start is the identification of 2 distinct categories of components: those focussed on raw data collection and those focussed on research and editing. The former category of functionality could be developed first and thus an early invitation could be issued to participate in the community. Developing in iterations is critical to the success of software development. A project should not seek to build its entire scope before the first release.¹⁷¹

The Virtual Manuscript Room Collaborative Research Environment (VMR CRE) is an entirely web-based solution which supports globally disparate teams as they edit a critical edition. It has facilities to encapsulate this work from start to finish:

- Cataloging manuscripts;
- Imaging manuscript pages: supports local links to images, remote links by dragging and dropping between browser windows from a remote site, and remote links using remote IIIF manifest files;

¹⁶⁹ Kate Harvey, "Feature Creep: What Causes It & How To Avoid It," Chargify, July 12, 2016, <https://www.chargify.com/blog/feature-creep/>.

¹⁷⁰ The NTVMR is the instance of the VMR CRE software in Münster for the work of the ECM.

¹⁷¹ boss, "Feature Creep and the Importance of Iterations," *The MIT Entrepreneurship Review* (blog), December 1, 2010, <http://miter.mit.edu/articlefeature-creep-and-importance-iterations/>.

- Indexing content on pages;
- Transcribing: WYSIWYM editor which produces regular well-formed EpiDoc-inspired TEI saved to a versioned transcription repository. Tools are included to merge work done by two independent transcribers-- a quality assurance practice used by the ECM project-- with a manager comparing and reconciling these to a final, published transcription;
- Regularizing variation in a transcription to a normal form;
- Dynamic collation (comparing) of witnesses, with the integration of CollateX and visual representations as an alignment table, variant graph, or traditional negative apparatus;
- Editing an apparatus;
- Publishing an online critical edition, providing facilities to show the edited apparatus or to dynamically re-collate desired combinations of witnesses and regularization rules, to transparently move from an apparatus entry to a transcription to an image of the page from where that transcription was rendered.

3.8.1. Use cases

Collecting raw data is the foundation to all other research in this field. This task of data collection has been broken down into 4 primary tasks:

- **Cataloguing manuscripts:** A catalog manager should have unique rights and capability to manage a catalog of manuscripts. Creating and editing new

manuscripts must be simple and handle the full feature set of the historic manuscript catalogs in this discipline, with the ability for the manager to easily extend the list of features recorded about a manuscript or manuscript page. Any registered user should have the ability to tag a manuscript or manuscript page with an extended set of manuscript feature or page feature tags.

- **Manuscript image management:** A microfilm scanner or other image manager should have unique rights and capability to easily upload, and make images available online-- without requiring the manager to know any of the complexities of image processing software. Image assignment to a manuscript should be done with simple drag and drop of an image onto the correct manuscript page. Selecting image view permissions must be possible. The ability to supply attribution credit for the images must be allowed and displayed when presenting the image to a user. Thumbnails should be generated by the system and image viewing capabilities made available to the end user with appropriate access.
- **Page Indexing:** Any registered user should be allow to add indexing data to a page where no data exists for that page. Any user should be able to update their own contributed indexing data. An index manager should have the capability to modify any indexing data. The system should accept and parse indexing data input supplied in intuitive formats easy for any user to formulate naturally, with no, or very limited training.

- **Page transcribing:** Any user should be allow to make personal transcriptions of any manuscript in the system. When a transcriber needs assistance with uncommon content found in a manuscript, a paleography database should be searchable by what content can be derived by the transcriber. A transcriber should be allow to select a region on an image and add new entries to the paleography database when new graphemes are encountered. Transcriptions should be versioned to record modification. Three levels of transcription should be allowed: personal, project, published. Transcriptions should be entered within a web-based WYSIWYM tool shielding the text scholar from the complexities of XML markup, assuring a regular use of TEI. A transcription manager should have the facility to show differences between two personal transcriptions, with easy reconciliation of differences between the two into a single project transcription. A transcription manager should be allowed to publish a project transcription to designate it ready for general availability.

After facilities to accommodate this raw data collection have been achieved, research tools against this data will then be targeted:

- **Browse and query:** Any user should have the ability to browse the manuscript catalog beginning with general categories for manuscripts. Any user should have the ability to perform a multifaceted search against the entire dataset, specifying any possible combination of terms, feature tags, and ranges. When performing a search, a users can choose to return one of three types of results

which match their query: manuscripts, individual pages, or the clipped region boxes from feature tags where they have been provided.

- **Text collation:** Any user should have the ability to create manuscript lists and perform a collation of any of these lists, showing the results as either a traditional apparatus, a variant graph, or an alignment table. When performing a collation, a user should have the ability to choose regularization rules, which base text to collate against if performing a base collation, the verse to be collated, how many verses should be included in the result (up to 5).
- **Project management:** A user designated as a project manager should be able to create new projects, assign users to their project, assign manuscripts to their project, limit the scope of their project to a specific range in the documents' content (e.g., Gospel of John), choose a task for their project (transcribing, indexing, etc.), then assign users to tasks. A full project overview should show the manager a quick snapshot of the progress of their project, and members of the project should see exactly which tasks they are assigned and the progress of their work. A project should have the option to allow members of the project to claim responsibility for a task. A project should be allowed the option for open membership, allowing any user to join the project without approval.
- **Test passage tools:** A project manager should have the capability to choose a set of test passages. A team member should have the capability to designate to what reading a manuscript attests at a given test passage.

- **Standard miscellaneous collaboration facilities:** Provided by 3rd party components: user forums with the ability to post links directly to manuscript pages or other objects in the system to aid in discussion, bar chart statistics for entire site, activity streams as an RSS news feed for the entire site so one can see what work is actively being accomplished. All the facilities of a mature portal software which has been integrated into the framework should be made available to a project: project calendars, wikis, shared documents, blogs, polls, easy web page editor, personal and project web pages, knowledgebase facilities, news and announcements, user and group management, rights and roles management, etc.

3.9. Referenceability

A core development value of the system is to provide persistent, direct linking to any object, annotation, or query. Some examples follow:

3.9.1. Direct object views by persistent URL

Manuscript P5, folio 2r

<http://ntvmr.uni-muenster.de/manuscript-workspace?docID=10005&pageID=30>

The screenshot displays the NTVMR Manuscript Workspace interface. The top navigation bar includes links for Welcome, Manuscript Workspace, NT Transcripts, NT Conjectures, Forum, Lists, Transcribing, Indexing, Help, NA28, About, Collation, and Reset. The main content area is divided into three panels:

- Left Panel (Table of Contents):** A table listing manuscript pages with columns for Page ID, Content, and Image. The table shows pages 10 (1r, C2), 20 (1v, C1), 30 (2r), 40 (2v), 50 (3r, C2), and 60 (3v, C1). The current page, 30 (2r), is highlighted.
- Center Panel (Image Viewer):** A large image of folio 2r, showing a dark, irregular shape (possibly a stain or a piece of tape) on a light background. The text "YRUS 2484." is visible at the top of the image.
- Right Panel (Transcription Display):** A table showing the transcription of folio 2r. The table has columns for Col 1, Folio 2r, and a reference column. The transcription is in Greek, with some words highlighted in red. The text is organized into lines, with line numbers 1 through 25 on the left and right margins.

Manuscript P5, Folio 2r

The first καὶ on that page. Notice the pan position, zoom level, and annotation coordinates provided in the URL:

<http://ntvmr.uni-muenster.de/community/modules/papyri/?zoom=300&left=-1993&top=-991&box=802.333/382.6666/92.6666/73/&site=INTF&image=10005/10158/30/10/15>



Image Viewer, Display from a Persistent URL

3.9.2. Queries by persistent URL

Every feature tagged as an inscriptio in Codex Sinaiticus:

<http://ntvmr.uni-muenster.de/manuscript-workspace?docID=20001&searchType=clips&featureCode=Inscriptio>

The screenshot shows the NTVMR Manuscript Workspace interface. The top navigation bar includes links for Welcome, Manuscript Workspace, NT Transcripts, NT Conjectures, Forum, Liste, Transcribing, Indexing, Help, NA28, About, Collation, and Reset. The sidebar on the left lists manuscripts with columns for Page Id, Content, and Image. The central image viewer displays a page from the manuscript, with a red box highlighting a specific feature. The right-hand pane shows the transcription of the page, with a red box highlighting a specific feature.

Page Id	Content	Image
40 (200r)	Matt inscriptio; Matt 1:1-25; Matt 2:1-5	
360 (217v)	Mark 1:1-35	
570 (228r)	Mark 16:2-8; Luke 1:1-18	
950 (247r)	John 1:1-38	
1230 (261r)	Rom 1:1-32	
1360 (267v)	Rom 16:18-27; 1Cor 1:1-18	

The central image viewer displays a page from the manuscript, with a red box highlighting a specific feature. The right-hand pane shows the transcription of the page, with a red box highlighting a specific feature.

Incriptio in Sinaiticus

A representative α from all manuscripts in the system dating between the II and VII century:

<http://ntvmr.uni-muenster.de/manuscript-workspace?dateMin=100&dateMax=699&searchType=clips&featureCode=Grapheme%3D%CE%B1>

The screenshot shows the NTVMR Manuscript Workspace interface. The top navigation bar includes links for Welcome, Manuscript Workspace, NT Transcripts, NT Conjectures, Forum, Liste, Transcribing, Indexing, Help, NA28, About, Collation, and Reset. The main content area is divided into three sections:

- Manuscripts:** A table listing manuscripts with columns for Doc ID, Page Id, Content, and Image.
- Image Viewer:** A large image of a manuscript page with a highlighted area.
- Transcription Display:** A detailed transcription of the highlighted text, showing the original text and its transcription.

Doc ID	Page Id	Content	Image
10001	10 (1r)	Matt inscriptio-1:11	
10002	10 (1v)	John 12:12-15	
10003	10 (1r)	Luke 7:36-45	
10004	10 (1r)	Luke 1:58-60; Luke 1:62-73	
10005	10 (1r, C2)	John 1:23-31	
10006	10 (100)	John 10:1-2; John 10:4-7; John 10:9-10	
10006	20 (104)	John 11:1-7	

The Image Viewer shows a large image of a manuscript page with a highlighted area. The Transcription Display shows the text of the highlighted area, with the original text and its transcription.

Folio 1r

Col 1

1 βίβλος γενεσικ τῷ χυ σὺ δαυὶδ [ισακ]
2 ἀβραμ ἀβραμ ἐγεννησεν τὸν [ισακ]
3 [ισακ] δὲ ἐγεννησεν [ισακ] [ισακ]
4 δὲ ἐγεννησεν τὸν [ισακ] [ισακ]
5 [ισακ] ἐλφους αὐτοῦ [ισακ] δὲ ἐγεννη-
6 σεν τὸν φαρὲς καὶ τὸν ζαρε ἐκ τῆς θά-
7 μαρ φαρὲς δὲ ἐγεννησεν τὸν εσραμ
8 εσ[ραμ] δὲ ἐγεννησεν τὸν [ισακ] [ισακ]
9 δὲ [ισακ] ἐγεννησεν τὸν ἀμμινάδαβ
10 ἀμ[μινάδαβ] δὲ ἐγεννησεν τὸν ναασάν
11 ναα[σάν] δὲ ἐγεννησεν τὸν σαλ[μὼν]
12 σαλ[μὼν] δὲ ἐγεννη[σεν] τὸν βοεζ [εκ]
13 τῆς ραχβ βοεζ δὲ ἐγεννησεν τὸν ι-
14 ωβὴδ ἐκ τῆς ραχβ [ισακ] δὲ ἐγεννη-
15 σεν τὸν ιεσσα [ισακ] δὲ ἐγεννησεν
16 τὸν δαυὶδ τὸν βασιλεῖα [δαν] δὲ ἐγεν-
17 νησεν τὸν σολομὼνα ἐκ τῆς σαρκεῖο σο-
18 λομων δὲ ἐγεννη[σεν] τὸν [ισακ] [ισακ] ροβ-
19 αμ δὲ ἐγεννησεν [ισακ] [ισακ] αβ[ελ] δὲ
20 ἐγεννησεν [ισακ] [ισακ] [ισακ] δὲ ἐγεν-
21 νησεν τὸν ιωσάφατ [ισακ] δὲ ἐγεν-
22 νησεν τὸν ιωσάφατ [ισακ] δὲ ἐγεν-
23 νησεν τὸν [ισακ] [ισακ] δὲ ἐγεννησεν
24 [ισακ]
25 [ισακ]

Grapheme Display from Persistent URL

Any manuscript which has any part of the Gospel of John and also contains an illumination:

<http://ntvmr.uni-muenster.de/manuscript-workspace?biblicalContent=John&searchType=pages&featureCode=Illuminations>

The screenshot displays the NTVMR Manuscript Workspace interface. The top navigation bar includes links for Welcome, Manuscript Workspace, NT Transcripts, NT Conjectures, Forum, Liste, Transcribing, Indexing, Help, NA28, About, Collation, and Reset. The main content area is divided into three sections:

- Manuscripts:** A table listing search results for John 1:1-5.
- Image Viewer:** A large image of a manuscript page (John 1:1-5) with a colorful illumination.
- Transcription Display:** A section for the transcription of the selected manuscript page.

The table in the Manuscripts section is as follows:

Doc ID	Page Id	Content	Image
20009	3570 (179r)	John 1:1; John 1:3-5	
30015	3200 (158v)	John 1:4-9	
30015	3210 (159r)	John 1:10-14	
30017	5490 (266r)	John 1:1	
30124	2600 (128r)	John 1:1-11	

The Image Viewer shows a manuscript page with a large, ornate initial 'I' in blue and red, followed by the text 'IHN' in gold. The text is in Greek, and the page is numbered '179r' in the top right corner.

The Transcription Display section shows the text 'IHN' in gold, followed by the text 'IHN' in blue and red, and the text 'IHN' in gold. The text is in Greek, and the page is numbered '179r' in the top right corner.

Multifaceted Search Query from Persistent URL

3.9.3. Data by persistent URL

The XML data from the web services API for the above query:

<http://ntvmr.uni-muenster.de/community/vmr/api/metadata/liste/search/?biblicalContent=John&searchType=pages&featureCode=Illuminations&detail=page>

```
<?xml version="1.0" encoding="UTF-8"?>
<manuscripts count="24" pagecount="35">
  <manuscript docID="20009" primaryName="09" gaNum="09" v1ln="" lang="g" userID="" groupID="">
    <originYear early="800" late="899">IX</originYear>
    <contentOverview>e</contentOverview>
    <leavesDescription>204</leavesDescription>
    <columns columnsMin="2">2</columns>
    <lineCount lineCountMin="19" lineCountMax="22">19+</lineCount>
    <height heightMin="285"/>
    <width widthMin="220"/>
    <shelfInstances/>
    <pages count="1">
      <page pageID="3570" shelfID="0" folio="179r" biblicalContent="John 1:1; John 1:3-5" indexContent="John 1:1; John 1:3-5" indexedBy="MF" sortOrder="03570">
        <images>
          <image thumbURL="http://ntmss.info/images/thumbs/20009/3570_49_20.jpg"
            webFriendlyURL="http://ntmss.info/images/webfriendly/20009/20009x03570XX_INTF.jpg" accessLevel="0" protectLevel="0" instID="49"
            surrID="20" userName="" uri="http://ntmss.info/images/webfriendly/20009/20009x03570XX_INTF.jpg" viewURL="http://ntvmr.uni-muenster.de/community/modules/papyri/?site=VMR&image=20009/3570/20">
              <attribution/>
            </image>
          </images>
        </page>
      </pages>
    </manuscript>
    <manuscript docID="30015" primaryName="15" gaNum="15" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30017" primaryName="17" gaNum="17" v1ln="" lang="g-l" userID="" groupID="">...</manuscript>
    <manuscript docID="30124" primaryName="124" gaNum="124" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30162" primaryName="162" gaNum="162" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30386" primaryName="386" gaNum="386" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30389" primaryName="389" gaNum="389" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30544" primaryName="544" gaNum="544" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30561" primaryName="561" gaNum="561" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30579" primaryName="579" gaNum="579" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30700" primaryName="700" gaNum="700" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="30706" primaryName="706" gaNum="706" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31008" primaryName="1008" gaNum="1008" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31084" primaryName="1084" gaNum="1084" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31094" primaryName="1094" gaNum="1094" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31113" primaryName="1113" gaNum="1113" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31359" primaryName="1359" gaNum="1359" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31463" primaryName="1463" gaNum="1463" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31574" primaryName="1574" gaNum="1574" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="31593" primaryName="1593" gaNum="1593" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="32400" primaryName="2400" gaNum="2400" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="32561" primaryName="2561" gaNum="2561" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="40131" primaryName="l 131" gaNum="l 131" v1ln="" lang="g" userID="" groupID="">...</manuscript>
    <manuscript docID="41692" primaryName="l 1692" gaNum="l 1692" v1ln="" lang="g" userID="" groupID="">...</manuscript>
  </manuscripts>
```

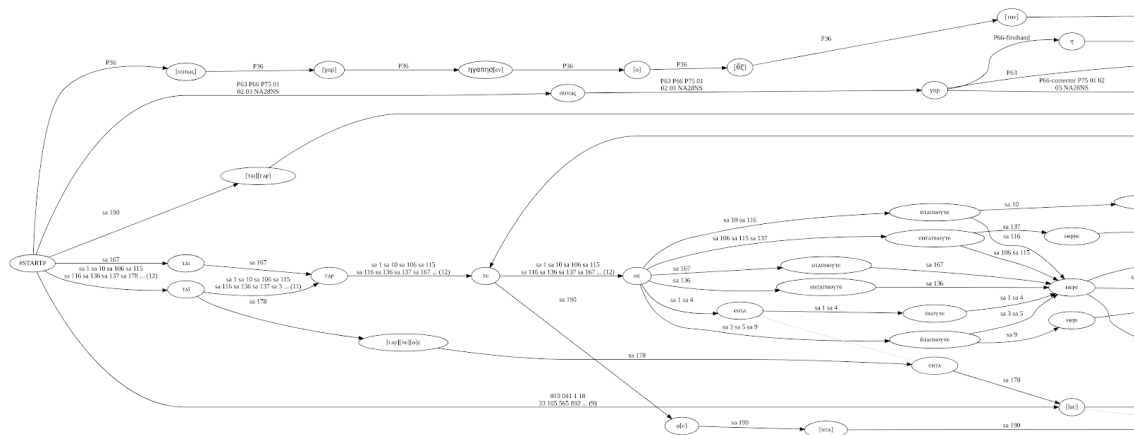
XML Output Format from a Persistent URL.

<http://ntvmr.uni-muenster.de/community/vmr/api/collate/?documentGroupID=-1&indexContent=jn.3.16>

3.9.5. Variant graphs from persistent URLs

Below is a persistent URL which collates all witness to John 3:16 in the system, unregularized, and displays the result as a variant graph:

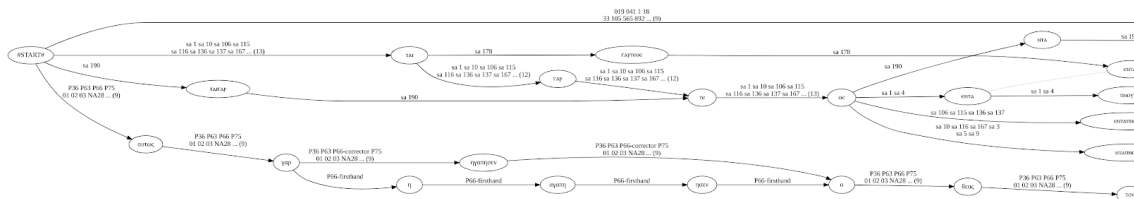
<http://ntvmr.uni-muenster.de/community/vmr/api/collate/?documentGroupID=-1&indexContent=in.3.16&format=graph>



Variant Graph from a Persistent URL

The same, including a base text and with regularization rules:

http://ntvmr.uni-muenster.de/community/vmr/api/collate/?baseText=NA28&verse=John.3.16&documentGroupID=-1&ignorePunctuation=true&ignoreSupplied=true&ignoreUnclear=true&format=graph®UserID=intfadmin*



Regularized Variant Graph from a Persistent URL

3.9.6. Workspaces which honor persistent URLs

An end user's reader view for the regularized collation of John 3:16:

<http://ntvmr.uni-muenster.de/nt-transcripts?verse=John.3.16>

The screenshot displays the 'NT Transcripts' workspace in a web browser. The interface includes a sidebar with navigation links: Welcome, Manuscript Workspace, NT Transcripts (highlighted), NT Conjectures, Forum, Liste, Transcribing, Indexing, Status, Help, NA28, About, Collation, and Reset. The main content area shows the text of John 3:16 in Greek, with a search bar at the top. Below the text, there is a 'Witness List' section with a table of manuscript references. A dropdown menu is open, showing options like 'Base Text', 'Regularization Rules', 'Which Transcriptions', 'Include Extra Verses', 'Show Alignment Table', 'Show Variant Graph', and 'Download CSV'. The right sidebar contains a 'Collation' section with a table of variant readings. The bottom of the page has a footer with 'Settings' and 'Online Friends (4)'.

WESTFÄLISCHE
WILHELM-UNIVERSITÄT
MÜNSTER

Institut für
Neutestamentliche
Textforschung

NA28 Search

Quick Lookup: John 3:16

John 3:16

14 και καθως μωυσης υψωσεν τον οφιν εν τη ερημω ουτως υψωθηναι δει τον υιον του ανθρωπου
15 ινα πας ο πιστευων εν αυτω εχη ζωην αιωνιον
16 ουτως γαρ ηγαπησεν ο θεος τον κοσμον ωστε τον υιον τον μονογενη εδωκεν ινα πας ο πιστευων εις αυτον μη αποληται αλλ εχη ζωην αιωνιον
17 ου γαρ απεστειλεν ο θεος τον υιον εις τον κοσμον ινα κρινη τον κοσμον αλλ ινα σωθη ο κοσμος δι αυτου
18 ο πιστευων εις αυτον ου κρινεται ο δε μη πιστευων ηδη κεκριται οτι μη πεπιστευκεν εις το ονομα του μονογενου υιου του θεου
19 αυτη δε εστιν η κρισις οτι το φως εληλυθεν εις τον κοσμον και ηγαπησαν οι ανθρωποι μαλλον το σκοτος η το φως ην γαρ αυτων πονηρα τα εργα

Not Present in P2 P5 P6 P22 P28 P39 P44 P45 P52 P54
P60 P76 P80 P84 P90 P93 P95 P106 P107
04 05 18 33 81 323 614 630 945 1241

Present in P36 P63 P66 P75 01 02 03

Base: NA28]

ηγαπησεν]
η P66-firsthand

τ]
αγαπηη P66-firsthand

τ]
ησεν P66-firsthand

τ]
αυτου P36 P63 01-corrector1 01-corrector2 02

Witness List

Base Text

Regularization Rules

Which Transcriptions

Include Extra Verses

Show Alignment Table

Show Variant Graph

Download CSV

ἀπόλλυμι

I. to destroy utterly, kill, slay, and of things, to destroy, demolish, waste, Hom., attic; in pregnant sense, γὰς ἐκ πατρίδος ἀπ. to drive me ruined from my fatherland, Eur.; λόγοις ἀπ. τινά Soph.—to talk or bore one to death, Ar.

II. to lose utterly, πατέρα, νόστιμον ἡμῶν Hom.

B. Mid. ἀπόλλυμαι, fut. -ολοῦμαι, ionic -ολέομαι with part. ἀπολεόμενος; aor2 -ωλόμην; perf. -όλωκα; plup. ἀπολώλεον— to perish utterly, die, il.; c. acc. cogn. ἀπ. κακὸν μέρος, ἀπὸν ὄλεθρον Od.; to be undone, id=Od.; ἀπόλωλας you are done for, lost, ruined, Ar.; as an imprecation, κόκισι' ἀπολόμην id=Ar.; in fut. part., ὦ κόκισι' ἀπολούμενε oh destined to a miserable end! i. e. oh villain, scoundrell Ar.

II. to be lost, slip away, vanish, of the water eluding Tantalus, Od.; of sleep, il.

from Liddell and Scott, An Intermediate Greek-English Lexicon
Courtesy of Perseus Digital Library

Settings Online Friends (4)

Reader View, Realtime Collation of John 3:16 from a Persistent URL

<http://ntvmr.uni-muenster.de/collation?verse=John.3.16&collate=graph>

Regularization Workspace at John 3.16 from a Persistent URL

3.10. Multilingual

The VMR CRE has been designed from its inception to support a wide array of languages. A demonstration of transcribing in Arabic, Armenian, Coptic, Ethiopic, Greek, Hebrew, Slavonic, and Syriac can be seen from this link:

<http://ntvmr.uni-muenster.de/web/test/comst>

Click the language of interest and then click the [from Basetext] button on the transcription editor to load the indexed verses into the editor as a starting point for transcribing in that language.

4. Stages of editing a critical edition

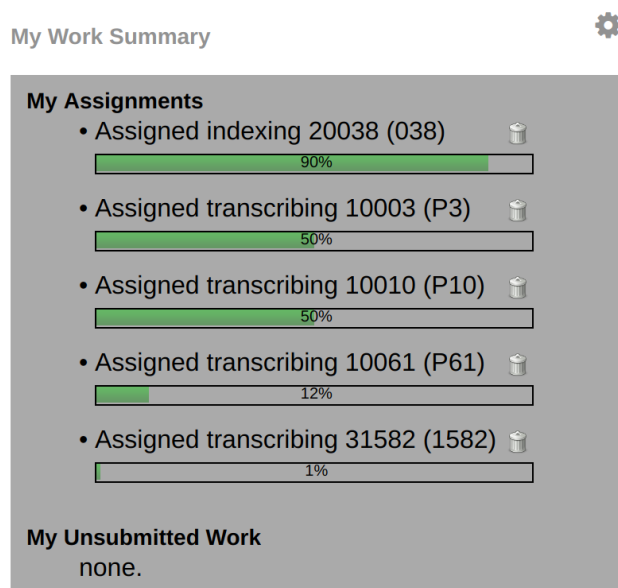
Editing a fascicle for the ECM has an established workflow at the INTF. It is not uncommon for multiple fascicles to be in parallel production, all at different stages within the editing process. When this research began in 2010, the Greek apparatus for Acts was nearly complete and CBGM analysis had begun. Also at the INTF, the transcriptions and reconciliation of those transcriptions was ongoing for Mark. The manuscripts for Matthew were being imaged and indexed. Along with this work at the INTF, the IGNTP in Birmingham was midway through transcribing and reconciling for the Gospel of John. Now, at the time of this writing in 2017, Acts is complete and days away from print availability, the Greek apparatus for Mark is midway complete, transcription and reconciliation for Matthew is well underway. In Birmingham, versional support is being added to the completed Greek apparatus of John. A new project to edit the Apocalypse at the ISBTF has completed the selection of manuscripts and continues in transcription and reconciliation. New projects are in their infancies at external institutions to edit Galatians and I Timothy. Each of these efforts are following roughly the same workflow, which can prove an informative template for other editing projects. The following sections outline this workflow¹⁷² to produce a fascicle of the ECM, delineating the work into 8 discrete stages for the editing process. In each section following, first the stage of editing work is discussed and afterward, the relevant

¹⁷² Summaries of the ECM workflow can also be found in Parker, *An Introduction*, and Houghton and Smith, “Digital Editing and the Greek NT”.

VMR CRE components developed to facilitate that stage of the editing process are presented.

4.0.1. My Work Assignments gadget


Before discussing the specialized components which cater to an individual stage of editing, first components presented here are gadgets generally useful to facilitate online collaboration, as a whole. A vital part of building a thriving online community involves encouraging and facilitating contribution. A willing contributor needs to know what work is available, have the opportunity to claim responsibility for a task or to be assigned work by a manager, see their assignments, the progress of their work, jump to the where they left off last, etc. The My Work Assignments gadget can be included on a workspace to show the contributor what task responsibilities have been assigned to them and also serves as a launchpoint to quickly jump back into a task.






My Work Summary Gadget


4.0.2. Project Status gadget

Many stages of work on a project need to be organized by a manager and assigned to a qualified individual. In [section 2.4](#), we discussed the need for a collaborative community project to grant easy entry into the work by allowing a user to claim responsibility for entry level tasks. This is not appropriate for many stages of work which require special skills. The Project Status gadget can be configured to disallow sub-project team members from claiming their own tasks, giving control to a manager to organize and assign work. Below is a view of the ECM Matthew transcription status.

Project Status 

ECM Matthew

Status	Assignment	DocID	Name	Date	Proj Pgs	Total Pgs
Manuscripts: 143						
		+ 20024	024	VI	18	44
 <div>0% Complete (View)</div> <div>0% Reconciled (View)</div> <div>100% Complete (View)</div>	jufer ×	+ 2003				160
	stedwards ×					
		+ 2003				32
		+ 2004				188
		+ 20043	043	VI	21	197
 <div>100% Complete (View)</div> <div>0% Reconciled (View)</div> <div>100% Complete (View)</div>	jufer ×	+ 20073	073 (+084)	VI	7	?
	lakmann ×					
 <div>100% Complete (View)</div> <div>0% Reconciled (View)</div> <div>100% Complete (View)</div>	jufer ×	+ 20085	085	VI	6	3
	lakmann ×					

Create New Assignment 

New Assignment for Document: 20024
Assign to Project Member:

[drubschmid](#) ▼

Create Assignment

Project Status Gadget, Disallowing Open Assignments¹⁷³

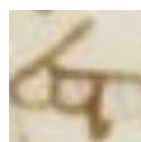
¹⁷³ For a view of the Project Status gadget configured to allow the community to select work, see [section 2.4](#)

4.1. Cataloguing Manuscripts

Before any other work can begin toward a critical edition of a text-- before deeper evaluation of each manuscript can begin, those witnesses which attest to the text must be catalogued. With ECM projects, this work was completed long ago and the catalog is updated with excitement when a New Testament manuscript discovery is made. This catalog at the INTF has long been recognized as the authority for known New Testament manuscripts.

4.1.1. The *Kurzgefaßte Liste*

As mentioned in [chapter 1](#), the *Kurzgefaßte Liste* catalog was initially established with the publication of the *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments* by Kurt Aland. The *Liste*, and thus the INTF, divides manuscripts into four categories: Papyri, Majuscules, Minuscules, and Lectionaries. This order follows the history of both the canvas and also Greek letter form. The Papyri are generally the oldest witnesses with the Majuscules following. Both are written in Majuscule script-- discrete, block uppercase Greek letters. Majuscules switched away from papyrus to parchment. The Minuscules display an evolved Greek scripts which includes both upper and lowercase connected characters, often replete with ligatures-- shapes which represent common letter combinations. See the ligature representing εστ below.



Lectionaries were used in church services and contain biblical readings which typically follow two liturgical reading schedules: Synaxarion based on festival periods, weeks of those periods, and the day of week, and Menologion which organize readings by month, starting in September, and day of month. Both types of schedule readings include singing tone, at which service the reading is to be used, the reading number for that service, intros (“prokeimenon”) and subscripts (“alleluia”), and which saint or holy day they commemorate.¹⁷⁴ Although not continuous text Greek New Testament manuscripts, lectionaries include a vast amount of biblical content and are useful witnesses for establishing the text of the New Testament.

Today, manuscripts used for the ECM are located all over the world and often an individual manuscript is divided between institutions. The *Liste* catalogue gives an identifier called a Gregory-Aland number or GA Number to each manuscript and includes the library shelf number information for all parts of each manuscript. All of this information and more need to be captured in a manuscript catalogue.

4.1.2. The concept of a manuscript identifier

One may think that the concept of a manuscript identifier is fairly straightforward. To understand some often overlooked nuances of difficulty, a short review will be helpful of two projects which now collaborate with the INTF.

¹⁷⁴ David C. Parker, *An Introduction to the New Testament Manuscripts and Their Texts* (Cambridge, UK: Cambridge University Press, 2008), 56.

4.1.2.1. The Pinakes Project

The Pontifical Institute of Mediaeval Studies, Toronto, Canada¹⁷⁵ began work on the Pinakes Project¹⁷⁶ in 1971, as an attempt to consolidate catalogs of Greek manuscripts into a single publication. In 1993 the resources and ownership were transferred to the *Institut de recherche et d'histoire des textes*, Paris, who published the first online version in 2008. As of August, 2017, Pinakes has a rich online catalog of more than 70,000 Greek object entries worldwide. Diktyon Numbers are used as the key identifiers for manuscript objects within the Pinakes database. These identifiers have been assigned sequentially, beginning at 1, to all shelf numbers which contain one or more Greek folios across all holding institutions included in the Pinakes database; Diktyon Number identifiers retain no classification or other meaning in their assignment. In short, a Diktyon Number is a universally unique identifier, across all holding institutions, for a single library shelf number (hereafter referred to as a “shelf instance”); conceptually, there is a one to one relationship between a Diktyon Numbers and a shelf instance. For example, Codex Vaticanus, which has a Gregory-Aland number of 03 or label ‘B’, has a shelf number at the Vatican Library of gr. 1209, and has this Pinakes Diktyon Number:

67840 - Vaticano Vaticano Biblioteca Apostolica Vaticana Vat. gr. 1209¹⁷⁷

¹⁷⁵ Pontifical Institute of Mediaeval Studies, “PIMS: Pontifical Institute of Mediaeval Studies,” PIMS, 2017, <http://www.pims.ca>.

¹⁷⁶ *Institut de recherche et d'histoire des textes*, “Pinakes - Textes et manuscrits grecs,” Pinakes, 2016, <http://pinakes.irht.cnrs.fr>.

¹⁷⁷ *Institut de recherche et d'histoire des textes*, “Biblioteca Apostolica Vaticana Vat. gr. 1209,” Pinakes, 2016, <http://pinakes.irht.cnrs.fr/notices/cote/67840/>

This simple definition of a manuscript identifier lends itself nicely to the work of Pinakes-- to consolidate catalogs, and avoids any scholarly debate about the contents within a shelf instance. This is not to say that consideration and theoretic discussion about manuscript objects and their content do not happen within the Pinakes project; only that the primary identifier, the Diktyon Number, itself, avoids this debate by simply denoting a library shelf instance. In addition to using Diktyon Numbers to identify manuscript objects, Pinakes also employs the concepts of a Work Identifier and Author Identifier. These are used to address the text within a manuscript. A Diktyon Number may be associated with one or more Work Identifiers and these associations carry data to describe the location of the instance or instances of that work within the manuscript.

4.1.2.2. The ParaTexBib Project

The ParaTexBib project¹⁷⁸ is based in München and under the direction of Martin Walraff, led by Patrick Andrist, and originally supported technically by Jenny Goudet who died in 2016. ParaTexBib has partnered with Pinakes and this present project at the INTF to identify and label paratextual material¹⁷⁹ within Greek New Testament manuscripts using the feature tagging facility (see [section 3.7](#)) of the software system. The data produced by the project primarily lives in the Pinakes database as additions to Work and Author Identifiers, and associations between Diktyon Number and paratextual Work Identifier to include the folio or page locations for the instance of the

¹⁷⁸ originally begun in Basel and relocated to München in 2016; ParaTexBib Project, "Paratexts of the Bible," ParaTexBib, 2014, <http://www.paratexbib.eu>.

¹⁷⁹ Paratextual material is most everything in a manuscript that is not biblical text, e.g., ancient chapter lists, introductory texts, apparatuses to aid the reading of the biblical text, etc.

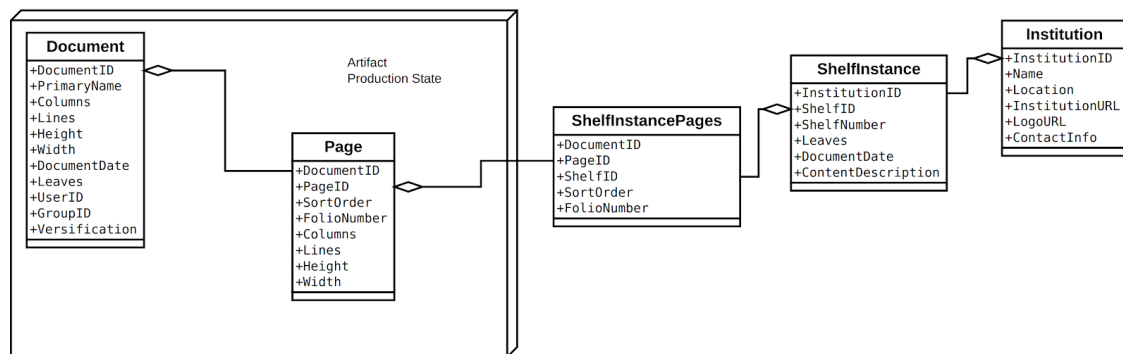
paratext within the manuscript. The ParaTexBib team also makes an attempt to label two additional concepts of a manuscript. First, they distinguish between “production units” (“*unité de production*”), which they define as initial manuscript production done by the same producers at roughly the same time. Different from production unit, they also attempt to identify what they have named “circulation units” (“*unité de circulation*”) of a manuscript. These would include one or more production units and designate a distinct artifact object which, at some point in time, was circulated to one or more recipients.¹⁸⁰

To facilitate this collaboration between Pinakes, the ParaTexBib Project, and the NTVMR, Diktyon Numbers have been added to the NTVMR as alternate identifiers for manuscripts. Here, an interesting challenge has arisen: The concept of a manuscript object between the NTVMR and Pinakes is different. In the NTVMR, manuscript objects are identified with Gregory-Aland numbers, which identify the New Testament portion within conceptually reconstituted ancient entities (largely equivalent to “circulation units”). In Pinakes, a manuscript identifier simply aligns to a library shelf instance at a holding institution, regardless of whether the box on that shelf contains leaves from multiple ancient objects, the partial leaves from a single object which is distributed across multiple holding institutions, or some other combination thereof. For example, the NTVMR gives Gregory-Aland number 01 or the label ‘ α ’ to Codex Sinaiticus. In Pinakes, this production unit has these Diktyon Numbers:

¹⁸⁰ Patrick Andrist, Paul Canart, and Marilena Maniaci, *La Syntaxe Du Codex: Essai de Codicologie Structurale*, Bibliologia 34 (Turnhout: Brepols, 2013).

- **39225** - United Kingdom London British Library Add 43725
- **38316** - Deutschland Leipzig UB gr. 01
- **57069** - Rossijskaja Federacija Sankt-Peterburg Rossijskaja Nacional'naja biblioteka (RNB) Ф. № 906 (Gr.) 002 (Granstrem 2)
- **57331** - Rossijskaja Federacija Sankt-Peterburg Rossijskaja Nacional'naja biblioteka (RNB) Ф. № 906 (Gr.) 259 (Granstrem 2)
- **57913** - Rossijskaja Federacija Sankt-Peterburg Rossijskaja Nacional'naja biblioteka (RNB) Ф. № 906 (Gr.) 843 (Granstrem 2)
- **57946** - Rossijskaja Federacija Sankt-Peterburg Rossijskaja Nacional'naja biblioteka (RNB) Ф. № 536. On. 1. Sobr. Obščectva Ljubitelej Drevnej Pis'mennosti 0 156 (Granstrem 3)
- **61039** - Egypte Sinai Monê tês Hagias Aikaterinês NE gr. MГ 001

The NTVMR manuscript identifier represents an attempt to conceptually reconstitute an ancient object and the Diktyon Number identifier does not. The relevant section of the VMR CRE data model follows.



Catalog Data Model

As can be seen, both concepts are present in the data model. The primary object identifier at the INTF is the DocumentID in the Document table. Documents have a one to many relationship with Pages, as one from the perspective of researching a

historical document would expect. This parallels the ParTexBib circulation unit.¹⁸¹

Also in the data model, from right to left in the image previous, Institutions have a one to many relationship with ShelfInstances-- one holding institute may have many boxes on their shelves which contain manuscript material. These ShelfInstances also have a one to many relationship with Pages-- each box on the shelf may contain material for more than one Page. The bridge between the ancient and the situation today-- the left vs. the right of the data model segment-- is the center ShelfInstancePages tables. Both sides have a one to many relationship with this table. This allows a ShelfInstance to have many Pages and it also allow a Page to be in many ShelfInstances-- in the case of fragmentary material. The sole purpose for the ShelfInstancePages table is to make relationships between ancient manuscript pages, and where they reside today at holding institutes. The issue still to be resolved is that the concept of an alternative catalog identifier in the VMR CRE has been relegated to manuscript feature tagging which can associate this metadata with a manuscript, but not a shelf instance. This is sensible for the field of New Testament textual criticism, as all other catalogs in this field also focus on, and thus key their catalog against, the concept of an ancient manuscript-- von Soden IDs, Tischendorf IDs. Many other catalogs outside this specialized research field follow suit-- Leuven Database of Ancient Books (LDAB),¹⁸² Trismegistos,¹⁸³ Oxford's

¹⁸¹ The pages of supplemental material which may have been added to a document over time, in the INTF *Liste* catalog are usually contained within a single document object with the pages marked as supplemental. In few cases, however, substantial supplements have received their own GA number, e.g. the 15th c. supplement of Codex Vaticanus which is labelled GA 1957.

¹⁸² Leuven Database of Ancient Books, "Leuven Database of Ancient Books," LDAB, 2017, <http://www.trismegistos.org/ldab/>.

¹⁸³ Trismegistos, "Trismegistos - An interdisciplinary portal of papyrological and epigraphical resources," Trismegistos, 2018, <http://www.trismegistos.org>.

Oxyrhynchus Papyri POxy¹⁸⁴ catalog. These other “alternative identifier” feature tags in the VMR CRE system benefit from the auxiliary data fields available for storing extra data along with a feature tag. They store not just the alternate catalog ID, but often a URL directly to the online catalog entry. This same mechanism has also resolved our difficulty of designating, for a Diktyon Number, a single shelf instance within a manuscript by allowing selection of the ShelfInstance object of a manuscript (GA number) to which the Diktyon Number relates.

The screenshot shows a web form titled "Feature Data" with a close button (X) in the top right corner. The form contains several input fields:

- Feature 123205:** A dropdown menu with "Diktyon Number" selected.
- URL:** A text input field containing "http://pinakes.irht.cnrs.fr/notices/cote/id/11950".
- Diktyon Number:** A text input field containing "11950".
- Shelf ID:** A dropdown menu with a list of options: "B. XVII. 20, 21; Trinity College; Cambridge, United Kingdom", "Cod. 91 in scrin.; Staats- und Universitätsbibliothek; Hamburg, Germany", and "B. XVII. 20, 21; Trinity College; Cambridge, United Kingdom". The last option is highlighted in green.
- Additional Notes:** A large text area for additional information.

At the bottom left of the form is a "Save" button with a green downward arrow icon.

Diktyon Number Alternative Identifier Feature Tag

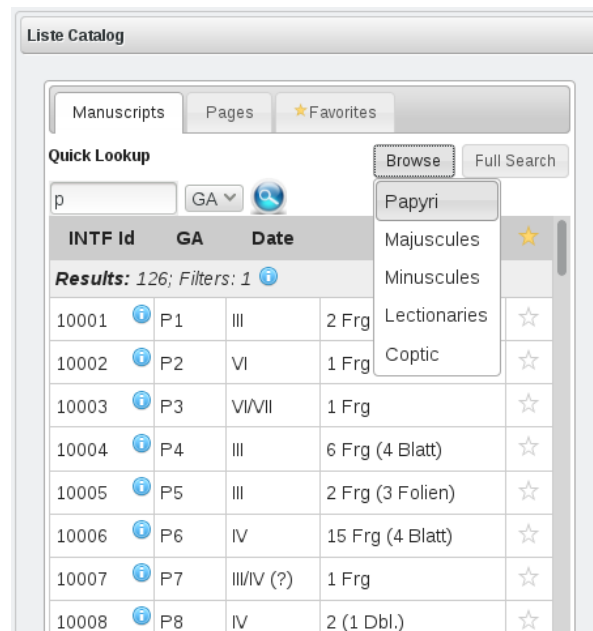
4.1.3. The Catalogue gadget

The digitization of the data within the printed *Liste* was a principal task preparing for the VMR 1.0 and succeeded in format shifting¹⁸⁵ this important data from paper to bits.

¹⁸⁴ Imaging Papyri Project, “POxy: Oxyrhynchus Online,” University of Oxford, 2017, <http://www.papyrology.ox.ac.uk/POxy>.

¹⁸⁵ Format shifting describes the conversion of data from one format to another-- typically between digital formats, but here describes a more important shift from paper to digital.

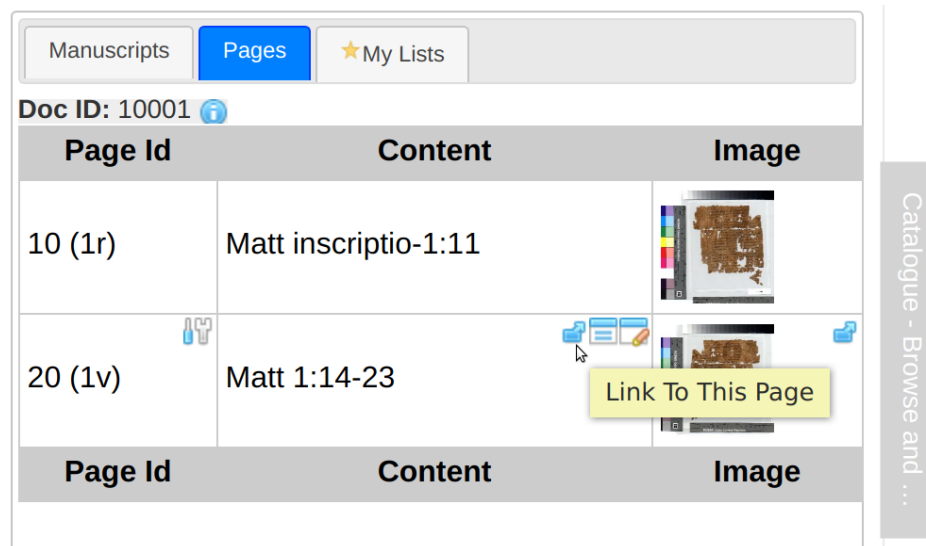
Now with a normalized data model and regularized data, a multifaceted query of this catalog can retrieve rich results.



Catalogue Gadget, Browsing

First, a **Quick Lookup** control in the upper left facilitates lookup of a known manuscript by Gregory-Aland Number or INTF ID. This control allows partial matches, as can be seen in the above image showing all manuscripts whose GA number begins with 'P' returned from the search performed. The **Browse** button allows easy population of the quick lookup control to show all manuscripts in a given category. At the top of the manuscript result list is shown the total number of documents which match the criteria, as well as how many and which filters were used to produce the list of manuscripts. Hovering over the **(i)** after the filter count will display search facet details. Each manuscript row lists the INTF ID, the Gregory-Aland number, the dating

assigned to the manuscript, the folio count details, along with a **Favorite star**, which can be used to mark a manuscript for easy retrieval later from the **My Lists** tab at the top. Hovering over the **(i)** after a manuscript's INTF ID will show more details, including holding institution information for the corresponding manuscript. Clicking on a manuscript row will load the pages of this manuscript and display them in the **Pages** tab, shown below.



Catalogue Gadget, Pages View

Again, the ability to see the manuscript details is offered by hovering over the info **(i)** icon following the INTF Doc ID. For each page, the INTF Page ID, the Folio Number, biblical content, and a thumbnail of the image is presented. As a reader hovers the mouse over a page row, the quick-tools toolbar is shown, seen as the five icons at the same horizontal level just above the cursor in the image above. These quick-tools buttons show up for the page nearest the mouse pointer and allow the researcher to edit the biblical content index information or to transcribe the page, to obtain direct links to

this same view of the page or to only the image for this page. If administrative access is granted to the user, the wrench and screwdriver icon will allow administrative editing functionality-- to relocate the page or a range of pages beginning here to another part of the manuscript, to publish a set of page transcriptions for this page or a range of pages starting at this point.

Beyond the basic lookup and navigation ability of the Catalogue gadget, the advanced multifaceted search controls can be found by clicking on the **Full Search** button in the upper right on the **Manuscript** tab, just after the **Browse** button. Selecting this feature will expose a facility which allows searching with any combination of components from the *Kurzgefasste Liste* resource to produce a desired set of manuscripts, manuscript pages, or a set of features from manuscript pages.

Full Search

Manuscript Num.
ID
Clear All

Current Country:

Place:

Institute:

Shelf Num.:

Indexed Biblical Content
John
3
16

Content Overview

Language
g

Has Feature

Dated To
100 - 499AD

Line Count
0 - ∞

Columns
0 - ∞

Page Width (mm)
0 - 450

Page Height (mm)
0 - 500

Has: Images
Transcriptions


Search
Show all:
docs
pages
feature clips

Catalogue Gadget, Full Search

The screen above shows a query for all Greek manuscript pages containing content from John chapter 3, verse 16, dated between the I and V century. Any of our metadata feature tags can be included by selecting the feature from the “Has Feature” dropdown, as shown below by the query for **feature clips** of all tagged Greek *alphas* before the VI century.

Full Search

Manuscript Num.
ID
▼

Clear All


Current Country:
▼

Place:
▼

Institute:
▼

Shelf Num.:
▼

Indexed Biblical Content
▼ ▼ ▼



Content Overview
▼



Language
▼



Has Feature
Grapheme
▼



Grapheme
=
▼
α
...



Has Feature
▼

Dated To
0 - 599AD



Line Count
0 - ∞



Columns
0 - ∞



Page Width (mm)
0 - 450



Page Height (mm)
0 - 500



Has: Images
☐
Transcriptions
☐

Search

Show all:
☐ docs
☐ pages
☒ feature clips

Catalogue Gadget, Feature Tag Search, Greek Alpha, pre-VI Century









First, notice that a second “Has Feature” row has been added. This allows more and more feature tag criteria to be added to a search. Also notice here the blue external link icon in the upper right corner, as seen throughout the VMR CRE, to generate a fully qualified link that can be used to return the reader or the reader’s audience, back to this specific context within the system. Here, the link represents this exactly query:

[http://ntvmr.uni-muenster.de/manuscript-workspace?dateMax=599&searchType=clips
&featureCode=Grapheme%3D%CE%B1&featureCodeClipString=Grapheme%3D%CE%B1](http://ntvmr.uni-muenster.de/manuscript-workspace?dateMax=599&searchType=clips&featureCode=Grapheme%3D%CE%B1&featureCodeClipString=Grapheme%3D%CE%B1)

The parameters included in the URL include:

- **dateMax** - the upper bound date to include,
- **searchType** - which type of results should be returned,
- **featureCode** - filter results by Grapheme= α , and
- **featureCodeClipString** - specifies which feature clips to show in the result list,
in the case where the search criteria includes multiple features.

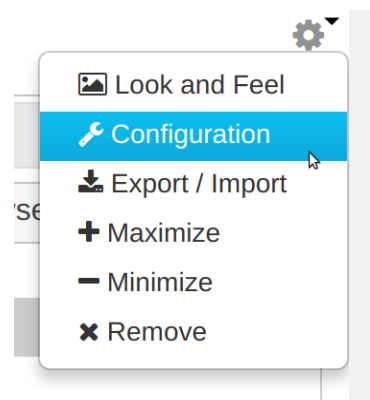
Clicking this link should return a reader to the system and perform this search, resulting in:

Doc ID	Page Id	Content	Image
10001	10 (1r)	Matt inscriptio-1:11	
10002	10 (1v)	John 12:12-15	
10003	10 (1r)	Luke 7:36-45	
10004	10 (1r)	Luke 1:58-60; Luke 1:62-73	
10005	10 (1r, C2)	John 1:23-31	
10006	10 (100)	John 10:1-2; John 10:4-7; John 10:9-10	
10006	20 (104)	John 11:1-7	
10007	10 (1r)	Luke 4:1-2	

Catalogue Gadget, Search Result, Feature Clips

4.1.3. The Catalogue gadget - configuration

As with all gadgets, the Catalogue gadget allows a set of custom configuration options which can be accessed when a user drops the gadget onto a page in their personal web space by clicking the settings gear and then the



configuration wrench at the top right corner of the gadget.

Catalogue - Browse and Search - Configuration ×

SetupPermissionsSharing

Archive/Restore Setup

Gadget Height

Short

Start In Full Search

No

Select First Search Result Automatically

No

Include Project Assignment Progress

No

Project Progress For

Current User

Include [Choose All] Button

Yes

Include Page Details

Yes

Include Indexing Details

No

Include Hover Info

Yes

Include External Image Viewer Quick Tool Button

Yes

Always Launch External Image Viewer

No

Always Launch External Transcription Viewer

No

Always Publish Image View Request

No

Append To Query String (e.g., docID=nnn)

Add Custom CSS resource to Gadget

Add Custom JS resource to Gadget

Save

Catalogue Gadget, Configuration

The **Gadget Height** selection allows the user to adjust the gadget's height. If **Start In Full Search** is set to 'Yes', the *Liste* Catalog gadget will begin with the **Full Search** control open. If the user wishes to always automatically make active the first search result, **Select First Search Result Automatically** can be set to 'Yes'. The Catalogue gadget can show the current user the progress of any project assignments they are viewing in the control. This is helpful if a user is, for example, transcribing a manuscript which they have been assigned. The gadget can show which pages have been assigned and which pages they have already completed. The **Include Project Assignment Progress** and **Project Progress For** settings can configure this functionality. Since gadgets are designed to be used in groups with other gadgets, the Catalogue gadget is sometimes useful for choosing which manuscripts or pages should be made active within a group of gadgets. For example, when collating a set of witnesses with the Collate gadget, the *Liste* Catalog gadget can be used to add a witness to the collation set. When used in tandem in this configuration, it is often helpful to have a **Choose All** button which can make active all results from a multifaceted search. This functionality can be included by selecting 'yes' for the **Include [Choose All] Button**. In other workspace designs, a researcher is using the Catalogue gadget to work exclusively with manuscripts, but not page details. **Include Page Details** can be set to 'no' to remove the **Pages** tab and functionality. **Include Indexing Details** can hide the page content index column if this is desired for a view. **Include Hover Info** can be used to hide the detail **(i)** icons and functionality if this is a hindrance for an

advanced user. The quick-tools button for showing an external image viewer can also be disabled with the **Include External Image Viewer Quicktool Button**. When selecting a manuscript page, often the details of that page can be seen in other gadgets present in the same workspace page as the *Liste* Catalog gadget; however, sometime a new popup window is instead desired for some surrogate representations of a page, e.g., image or transcription. The **Always Launch External Image Viewer** and **Always Launch External Transcription Viewer** can be use to force these surrogates into a separate window when a manuscript page is selected. When **Always Launch External Viewer** is set to true, any messages to request update of any image viewing gadgets on the same page are not published. To also force same page updates, even with external viewers turned on, set **Always Publish Image View Request** to ‘yes’. The **Append To Query String** allows query parameters to always be present when a user first arrives to a page with this instance of the Catalogue gadget. It can be used to always perform an initial search, as for a page designated to serve a palaeography course, or to select a specific manuscript for a page designed to highlight research featuring one witness, or to show a set of manuscripts, as for a course studying papyri. All query options are available to use here, to produce an initial view desired for a reader. The **Custom CSS** and **Custom JS** options open up the possibility for a web designer or programmer to affect the look and feel of the gadget.

The Catalogue gadget brings the full dataset of the *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments* into a completely searchable catalogue facility. It

has been useful internally to the staff at the INTF to confirm or deny new catalog submissions-- searching for all similarly described manuscripts to confirm the newly submitted object has not already been catalogued into the system. It is useful for quick navigation to an object for viewing, during transcription work, or most other tasks requiring manuscript and page navigation; indeed, it is rare to find a workspace layout within the VMR CRE which does not include the Catalogue gadget as a source for navigation-- hence, its many configuration options to allow it to be catered finely to its surroundings.

4.2. Choosing witnesses

When a large number of witnesses to a text exists-- many with identical readings at variant locations-- it becomes necessary to select a smaller, more practical set which is representative of the evidence. This is true not only to make the apparatus more readable, but also to limit the quantity of material to be exhaustively studied. Each project for the *Editio Critica Maior* essentially begins by reviewing (if they already exist), or repeating the process used to produce, the *Text und Textwert* volumes published from the beginning of the history of the INTF.¹⁸⁶ First, the witness set is narrowed down to only those manuscripts which have some portion extant of the section of the work the project seeks to edit. For example, from the approximately 5700 known Greek manuscripts which witness to the New Testament, only 552

¹⁸⁶ For the IGNTP a comparable method was developed in order to arrive at a smaller representative sample of manuscripts. See Frederik Wisse, *The Profile Method for the Classification and Evaluation of Manuscript Evidence, as Applied to the Continuous Greek Text of the Gospel of Luke*, Studies and Documents, v. 44 (Grand Rapids, Mich: Eerdmans, 1982).

(excluding lectionaries) contain some portion of the Catholic Epistles.¹⁸⁷ These witnesses can then be further pruned by utilizing “carefully selected test passages (*Teststellen*) for each book. By collating a manuscript only in a relatively small number of readings, rather than having to do a complete collation, it will be possible to know the quality of every NT Greek manuscript, at least as determined by the test passages.”

¹⁸⁸ For the Catholic Epistles, 98 *Teststellen* were used.¹⁸⁹ For Matthew and Luke, to “be successful in its aims and to meet the inevitable restraints of resources, the process of identification requires both the selection of satisfactory test passages and economy in the number chosen. As to the latter need, there are 64 *Teststellen* for Matthew and 54 for Luke. This works out as just over 2 per chapter, though they are not spread as evenly as that.”¹⁹⁰ This range is rather thin compared to the 196 test passages used for Mark.¹⁹¹ Speaking about the IGNTP work on editing John, Parker continues:

“In the case of the International Greek New Testament edition of John as part of the *Novum Testamentum Graecum Editio critica maior*, ... the starting point

¹⁸⁷ Larry W. Richard, “A Closer Look: Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments: Die Katholischen Briefe,” *Andrews University Seminary Studies* 34, no. 1 (Spring 1995). 37-46

¹⁸⁸ *ibid.*

¹⁸⁹ *ibid.*

¹⁹⁰ D. C. Parker, “Review of: Kurt Aland, Barbara Aland, and Klaus Wachtel, Eds., in Collaboration with Klaus Witte, Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments, Vol. 4: Die Synoptischen Evangelien, No. 2: Das Matthäusevangelium, and No. 3: Das Lukasevangelium,” *TC: A Journal of Biblical Textual Criticism* 5 (2000), <http://rosetta.reltech.org/TC/v05/Aland-et-al-ed2000rev.html>.

¹⁹¹ J. K. Elliott, “Review of: Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments, IV Die Synoptischen Evangelien, 2 Das Matthäusevangelium by Kurt Aland, Barbara Aland, Klaus Wachtel, Klaus Witte; Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments, IV Die Synoptischen Evangelien, 3 Das Lukasevangelium by Kurt Aland, Barbara Aland, Klaus Wachtel, Klaus Witte,” *Novum Testamentum* 41, no. 4 (1999): 390–92.


was the creation of a list of manuscripts containing some part of the Gospel of John according the Münster *Liste*... These manuscripts were scrutinised in detail in order to select the appropriate number to fulfil the *ECM*'s goal, to document 'the textual history of the first millennium.' ... First, as the basis for the selection of minuscule manuscripts, 153 test passages were selected in the first ten chapters, in which it would be possible to ascertain the extent to which witnesses supported the Nestle-Aland text and the majority text... The traditional cut-off point for selecting witnesses to be included in the edition, agreement with the majority text at 90% or fewer test passages, produced a shortlist of 301 manuscripts."¹⁹²


Because the *Text und Textwert* series did not include the Apocalypse, the team at the ISBTF in Wuppertal was required to first perform this process as they set out to select their manuscript list. The VMR CRE software module for this editing stage was designed and built in collaboration with the ISBTF team in their work on the *Apokalypse* volume of the *ECM*. The team defined 180 *Teststellen* to use while reviewing the manuscripts which have any portion of the Apocalypse extant.¹⁹³ The setup for the project began by designing and building an administrative tool to allow





¹⁹² David C. Parker et al., "The Selection of Greek Manuscripts to Be Included in the International Greek New Testament Project's Edition of John in the *Editio Critica Maior*," in *Studies on the Text of the New Testament and Early Christianity*, ed. Daniel M. Gurtner, Juan Hernández Jr., and Paul Foster (Leiden: Brill, 2015), 287–328, doi:10.1163/9789004300026_015.


¹⁹³ The final segment list chosen for the Apocalypse work can be retrieved using the following API URL: <http://ntvmr.uni-muenster.de/community/vmr/api/variant/list/?groupID=1&detail=segment>.

the project manager to enter and maintain this list of test passages. The Test Passages Admin gadget is shown below.

Test Passages - Admin 

 New Test Passage


ID	Verse	Context	Readings (id/text/wit. count)		Actions	
1	Rev 1:2	οσα ειδεν ... 3 μακαριος ο	1	---	x 136	 
			3	και ατινα εισι και ατινα χρη γενεσθαι μετα ταυτα	x 50	
			2	και ατινα εισι και α χρη γενεσθαι μετα ταυτα	x 49	
			1240	και ατινα εισι και ατινα χρη γενεσθαι	x 7	
			1247	και ατινα εισι και δια τινα χρη γενεσθαι μετα ταυτα	x 5	
			805	και ατινα εισι και χρη γενεσθαι μετα ταυτα	x 4	
			812	και	x 2	
			1535	και ατινα εστι και ατινα χρη γενεσθαι μετα ταυτα	x 2	
2	Rev 1:4	ειρηνη απο ... ο ων	5	θεου	x 137	 
			4	---	x 99	
			835	του	x 23	
			2045	ΟΜ. και ειρηνη απο *-* ο ων και ο ην και ο ερχομενος	x 1	
			1159	θεου πατρος	x 1	
			505	κυριου	x 1	

Edit Test Passage 

Verse

Test Passage Name

Context Description

Save 

Test Passage Administration Gadget

After the test passages were entered into the system, individual manuscripts were then assigned to team members for review, with the task to record the reading found at each *Teststelle* for their manuscript.¹⁹⁴ The gadget developed to facilitate this process for the Wuppertal *Apokalypse* project, the Test Passage Reading Selection gadget, allows a

¹⁹⁴ *Teststellen* are often checked from microfilm outside the system because many manuscripts are not yet imaged. After witnesses are determined, this sets priority for which manuscripts will next be digitized at the INTF.

researcher to choose the reading at each *Teststelle* for a given manuscript. The tool facilitates multiple hands, allows for an 'unsure' flag, and also provides for entry of the exact manuscript reading, along with editorial comments for each reading.

Test Passage Reading Selection

Document ID: 20001

Document ID: 20001 (01)

Verse: Rev 1:2 (SegID: 1)

οσα ειδεν ... 3 μακαριος ο

Reading: --- New

☐ V

Comment

► more hands...

Verse: Rev 1:4 (SegID: 2)

ειρηνη απο ... ο ων

Reading: --- New

☐ V

Comment

Hand: C2 New

▼ less...

Test Passage Reading Selection Gadget

These tools successfully enabled the ISBTF *Apocalypse* team to collaborate on the completion of the entire *Teststellen* dataset for their project in just over one year. The entire dataset can be retrieved for evaluation here:

<http://ntvmr.uni-muenster.de/community/vmr/api/variant/list/?groupID=1>

The *Text und Textwert* volume on the Apocalypse, based on the data collected through this interface, was published in 2017:

Markus Lembke et al., eds., *Die Apokalypse: Teststellenkollation Und Auswertungen*, Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments, herausgegeben von Markus Lembke, Darius Müller, Ulrich B. Schmid; 6 (Berlin: De Gruyter, 2017).

The completed *Teststellen* data allows statistical comparison of manuscripts to help determine groupings which inform the selection of a pruned manuscript list still representative of the transmission history. The gadget designed for statistical analysis of the similarity between manuscripts is called the CBGM Pre-Statistic gadget discussed more in [section 4.8.1.1](#). After an editorial team selects a witness list, digitizing images for those manuscripts becomes the next editing task.

4.3. Imaging resources

The Apocalypse was unique in that it was the only remaining book of the New Testament without a *Text und Textwert* volume completed before the development of the software system. It also had a relatively low number of extant witnesses-- about 300. This gave opportunity to digitize most microfilm for the edition before the selection of witnesses. More often, an ECM project will make their selection from a completed *Text und Textwert* volume and proceed to prioritize image digitization of those witnesses before each witness is indexed and transcribed. This priority is vital to the effective use of resources, seen clearly in the case of the Gospels, where typically ~1500 witnesses are available.

Providing image resources to manuscript researchers is a foundational feature for a collaborative manuscript research environment. Many obstacles, both political and technical, need to be hurdled with this task. First, permission needs to be obtained from the holding institute (and sometimes the photographer) to allow images to be made available to researchers within the software system. Next, determining if permission has been granted to store the image locally can be a challenge. Many institutes may already publicly post images for a manuscript from their institute's website and may also allow deep linking¹⁹⁵ to their existing images-- both a requirement for including external images into the VMR CRE software system with all the same capabilities as internal images. Ideally a copy of the images can be stored locally to the software system, which will prevent any breakage if an institution decides to move their images to a new URL, but this may not be allowed by a holding institution. Clarifying specifically and in writing from an image owner what permission is being granted is paramount to avoiding possible conflicts later. The VMR CRE supports both deep linking to external images and also local hosting, serving of image files from the VMR CRE server. After permissions are resolved, the technical tasks of hosting the images, linking these images to appropriate manuscript pages, generating web-friendly formats for display and thumbnails, archiving, and presenting to the researcher are all handled by the VMR CRE Page and Image Management gadget and the Image Viewer gadget. The VMR CRE divides image management and features into 5 primary tasks:

¹⁹⁵ Deep linking is a term used to refer not simply to a link which navigates a user to a resource at a different location, but instead pulls that remote resource for display within the current context.

- image storage,
- linking images to manuscript pages,
- setting viewing permissions and attributing credit for each image,
- generating appropriate viewing formats for each image,
- image viewing.

4.3.1. Image storage

The VMR CRE can manage and serve images. When first configuring the software system, the system's administrator must identify an image root folder on the VMR CRE server where images will be stored. The path then must be set within the **sysconfig.properties** configuration file using the two keys: `ImagesRootFolder` and `ImagesRootURL`. The former defines the full path to the chosen folder on the server. The latter defines the URL the web server hosting the VMR CRE is configured to serve this folder. For example:

```
ImagesRootFolder=/data/images/  
ImagesRootURL=http://ntmss.info/images/
```

An Apache httpd configuration might look like:

```
<VirtualHost *:80>  
    ServerName ntmss.info  
    Alias /images /data/images  
</VirtualHost>
```

```
<Directory "/data/images/">
    AllowOverride AuthConfig
    Require all granted
</Directory>
```

The images folder on the server should have 4 subfolders used by the software:

- full/ storage for full resolution, archived images,
- webfriendly/ location for generated .jpg images with a reasonable resolution for web research,
- thumbs/ location for thumbnails of each image (autogenerated),
- clips/ location for clips of features within an individual manuscript page (autogenerated).

Archiving is important for posterity and thus raw, high resolution scans should be kept isolated within the **full/** subfolder. This folder should be backed up regularly and archived off site for redundant protection. Images a project wishes to publish within the VMR CRE should be generated from the **full/** high resolution images into the **webfriendly/** folder. A reasonable web-appropriate resolution of no greater than 10MB JPEG format per image should be used. By convention, it is best to upload all images related to a specific manuscript into a subfolder named the same as the Document ID assigned to the manuscript which the images represent. This is not required within the

software system but makes organization easier and many of the image assistant tools will take advantage of this naming if it exists-- will provide extra assistance by initially scrolling an image file browser to an appropriate folder if found. The **thumbs/** and **clips/** subfolders should be created but the VMR CRE will manage files in these folders, automatically generating images when appropriate. Image names can be anything desired by the site administrators, within the realm of appropriate URL characters as defined by The Internet Engineering Task Force (IETF) RFC 3896, Section 2: Characters,¹⁹⁶ and not including reserved URL characters; in short, any characters within the set:

[ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789-._~!\$'()* ,]

Importantly, no spaces are allowed. As an example, a site might have a folder and file structure similar to:

/data/images/full/

/data/images/full/622031/

/data/images/full/622031/LBL_Or5000x02820xx.tif

/data/images/full/622031/LBL_Or5000x02690xx.tif

/data/images/full/622031/LBL_Or5000x02280xx.tif

/data/images/full/622032/

/data/images/full/622032/NBN_cass.B.19.488_folio098r_x_1_cmcl.tif

/data/images/full/622032/NBN_cass.B.19.488_folio098v_x_2_cmcl.tif

¹⁹⁶ The Internet Society, "RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax," The Internet Engineering Task Force, January 2005, <http://tools.ietf.org/html/rfc3986#section-2>.

/data/images/full/622032/NBN_cass.B.19.488_folio099r_x_3_cmcl.tif
/data/images/webfriendly/
/data/images/webfriendly/622031/
/data/images/webfriendly/622031/LBL_Or5000x02820xx.jpg
/data/images/webfriendly/622031/LBL_Or5000x02690xx.jpg
/data/images/webfriendly/622031/LBL_Or5000x02280xx.jpg
/data/images/webfriendly/622032/
/data/images/webfriendly/622032/NBN_cass.19.488_folio098r_x_1_cmcl.jpg
/data/images/webfriendly/622032/NBN_cass.19.488_folio098v_x_2_cmcl.jpg
/data/images/webfriendly/622032/NBN_cass.19.488_folio099r_x_3_cmcl.jpg
/data/images/thumbs/
/data/images/clips/

4.3.2. The Page and Image Management gadget

The Page and Image Management gadget in the VMR CRE allows a project administrator to define the number of pages (folio sides) which make up a manuscript, assign images to these pages, set viewing permissions and copyright attribution, and to generate the necessary derived image formats suitable for viewing in the VMR CRE.

The Page and Image Management gadget is often paired with the Institution Editor gadget and Shelf Instance gadget as seen in the image following.

Westfälische
Wilhelms-Universität
Münster

Institut für
Neutestamentliche
Textforschung

VMR Administration

Document Maintenance

Page and Image

Transcription Import

Feature Definitions

Birmingham Transcription Import

Project Management

VMR Administration / Page and Image

Image Folder: /data/images/webfriendly/10046/

Doc ID 10046

Document ID: 10046 (P46)

Hide Assigned

10042

10043

10044

10045

10046

raw

10046_001_INTF.jpg

10046_002_INTF.jpg

10046_003_INTF.jpg

10046_004_INTF.jpg

10046_005_INTF.jpg

10046_006_INTF.jpg

10046_007_INTF.jpg

10046_008_INTF.jpg

10046_009_INTF.jpg

10046_010_INTF.jpg

10046_011_INTF.jpg

10046_012_INTF.jpg

10046_013_INTF.jpg

10046_014_INTF.jpg

10046_015_INTF.jpg

10046_016_INTF.jpg

10046_017_INTF.jpg

10046_018_INTF.jpg

10046_019_INTF.jpg

10046_020_INTF.jpg

10046_021_INTF.jpg

10046_022_INTF.jpg

Apply

Clear Pending

Delete

Generate Thumbnails

10 (008r)

10046_001_INTF.jpg

T0005249-2.jpg

20 (008v)

10046_002_INTF.jpg

T0005250-1.jpg

30 (011r)

10046_003_INTF.jpg

T0004419.jpg

40 (011v)

10046_004_INTF.jpg

T0004418.jpg

50 (012r)

10046_005_INTF.jpg

T0004420.jpg

60 (012v)

10046_006_INTF.jpg

T0004421.jpg

70 (013r)

10046_007_INTF.jpg

T0004433-2.jpg

80 (013v)

10046_008_INTF.jpg

T0004432-1.jpg

90 (014r)

10046_009_INTF.jpg

D0005418-2.jpg

100 (014v)

10046_010_INTF.jpg

Tools

Select

Image Permissions

Generate Pages

External URL With Image Links

External URL Pattern

Local Files, RegEx Match

Folio Number From Image Name

Institution Editor

Country

Place

Institution Name

Admin - Shelf Instances

Shelf Instances for Manuscript: 10046. 0 Queued Pages

Institution	Shelf Number	Page Count	Actions
Chester Beatty Library Dublin Ireland	CBL BP II	112	
University Library Ann Arbor/Mich. USA	Inv. Nr. 6238	59	

Page and Image Management Workspace

4.3.2.1. Pairing Images with Manuscript Pages

The Page and Image Management gadget has a server-side file browser on the left side of the gadget which allows browsing folders and images uploaded to the server under the **webfriendly/** subfolder. These image files can be associated with manuscript pages by dragging and dropping from the left browser pane to manuscript pages in the center pane. This association can be done in bulk using regex¹⁹⁷ pattern matching with the

¹⁹⁷ Regular Expressions or regex is a pattern matching syntax used in many software products. To begin learning regex syntax, a helpful tutorial is available here: RegexOne, “RegexOne - Learn Regular Expressions,” RegexOne, 2017, <https://regexone.com>.

tools provided on the right. Before the Page and Image Management gadget can browse files on the server, proper configuration must be set in the system's **sysconfig.properties** configuration file to point to the top folder under which folders of images will be uploaded. To begin managing a manuscript, first the Document ID must be entered in the upper-right input box and loaded by hitting enter. If the manuscript is found, the button to the right of the input box will show the manuscript's friendly name. After loading, pages need to be added to the document using the "Generate Pages" tool on the right. By default, pages begin numbering at 10 and increment by 10 each page. This is an historical practice at the INTF which allows insertion of a missed scan or a new leaf discovered for a manuscript. The numbering itself does not carry any significance or other requirement than that it must be unique. Scholars are able to add desired folio numbers (1r, 1v, 2r, 2v, ...) in the indexing editorial stage described in [section 4.4](#). Once pages have been added with the "Generate Pages" tool, images can then be associated to these pages by dragging from the image file browser on the left to the appropriate page. Initially, work done adding pages and associating images will appear slightly transparent in the center panel. This signifies what work is still unsaved-- only staged. Once a set of work is satisfactory to the user, hitting the [Apply] button will save the staged work to the repository. This manual association of image files to created pages can become tedious with hundreds of images per manuscript. The remaining tools on the right of the Page and Image Management gadget ease this work by allowing the use of Regular Expressions and other assistance for bulk additions,

provided image files have been named uniformly and include pertinent information which can be extracted from the file name, including page number or folio numbers.

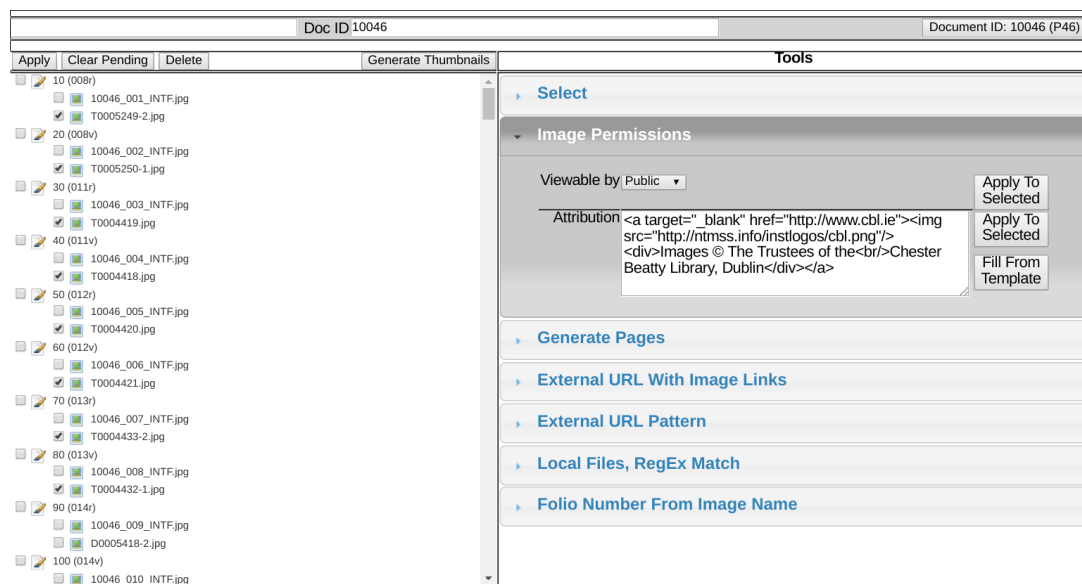
In addition to storing images on the VMR CRE server, publicly available images from other websites can also be associated with manuscript pages. This association is done by dragging the image from one browser window viewing the external website, onto another browser windows viewing the Page and Image Management gadget, in the same way local images are dragged from the left panel server image browser. These image will appear in the VMR CRE in the same way as if they were stored locally but the image will be loaded from the remote site each time a reader requests to view the manuscript image. Remote images must remain accessible at the same URL or dead links will prevent further access. Care must be taken to assure proper permission has been obtained from the publishing institution to allow users to access these remote images within the VMR and that proper credit is attributed to the publishing institution. Details for setting permissions and including attribution follow.

4.3.2.2. Viewing permissions and image accreditation

Once images are uploaded to the software system and these images-- or remotely published images-- are associated with manuscript pages for viewing by end users, viewing permission can be assigned and image credit can be attributed to each associated image. First, select the desired set of images and choose an access level from the following choices, then click the “Apply To Selected” button:

- **Public** - Any user can view the images.

- **Expert** - Only users assigned the “Expert User” role are able to view the images. This permission level and associated role are useful when an institution grants permission for images to be used for research purposes, but not publicly available. The software system does not automatically grant a login account the “Expert User” role.
- **Internal** - Only internal members of the project assigned the role “Internal User” will be able to view the images. The INTF uses this permission level and role for images which have been scanned from copyrighted facsimile editions or for other images which the institute has no permission to make available to anyone outside of the institution.



Page and Image Management Gadget, Setting Permissions and Attribution

Next, giving credit to the photographer or holding institution in exchange for granting permission to use images is crucial for maintaining a working relationship, important

when additional images might be desired in the future, and is generally an all around nice thing to do. The Image Permissions Attribution field can accept plain text and HTML markup, allowing an institution logo and link to be included with the attribution, if desired. Attribution credit is assigned to images in the same way viewing permission is assigned: first by filling in the attribution box, then selecting the desired images, and then pressing the “Apply To Selected” button.

4.3.2.3. Generating appropriate viewing formats

A future goal for the VMR CRE is to manage the **full/** resolution image subfolder and then to have the **webfriendly/** set of images automatically generated. This will provide the ability to configure a chosen, site-wide, web-friendly target resolution, and to automatically generate the **webfriendly/** subfolders and images within those subfolders according to the selected settings. In the future, a new resolution could then be chosen for the site and regeneration of the **webfriendly/** images at a new resolution could be automated. This is not currently the case. In the current release of the framework, the **webfriendly/** images must still be created manually by the site administrator. Many free utilities can process a set of archive images (likely .tif) and produce a desired resolution .jpg or .png set of files. Server-side scripts with ImageMagick¹⁹⁸ tools are utilized at the INTF to help batch process this procedure for a manuscript, but also some workers who digitize have local tools on their own computer to save an entire set of images for a manuscript at a web-friendly resolution. The VMR CRE does, however, provide the ability to generate the **thumbnails/** image set from the

¹⁹⁸ ImageMagick Studio LLC, “Convert, Edit, Or Compose Bitmap Images @ ImageMagick,” ImageMagick, 2017, <http://www.imagemagick.org/>.

webfriendly/ images, once assignment of images to manuscript pages has been made and saved. This is performed by pressing the “Generate Thumbnails” button at the top of the center pane of the Page and Image Management gadget. This process will automatically create the thumbnails shown to users while browsing through pages of a manuscript.

After creating pages and assigning images to those pages, setting permissions and attribution, then generating the thumbnails, the manuscript images are now available for viewing by the reader.

4.3.3. Image viewing

The VMR CRE image viewer is a ‘cleanroom’ (no external library references) implementation of an HTML / JavaScript image viewer for manuscripts, modeled after the ubiquitous Google Maps user experience. The viewer allows panning and zooming with the same mouse gestures as Google’s implementation, and also allows linkback to state via a dynamically updated anchor link. This is important, to allow scholars to have conversations about the objects they study in the viewer. As they view an image, zoom and pan to a selected region, and even annotate that region by dragging a box around an element of interest, they can then use the linkback anchor to obtain a URL which will bring others back to their view. This URL can be posted in forums or blogs or sent in an email, promoting conversation, in context, about their research. The

Image Viewer gadget can display local and remote static and IIIF¹⁹⁹ image resources.

The gadget can be seen below:

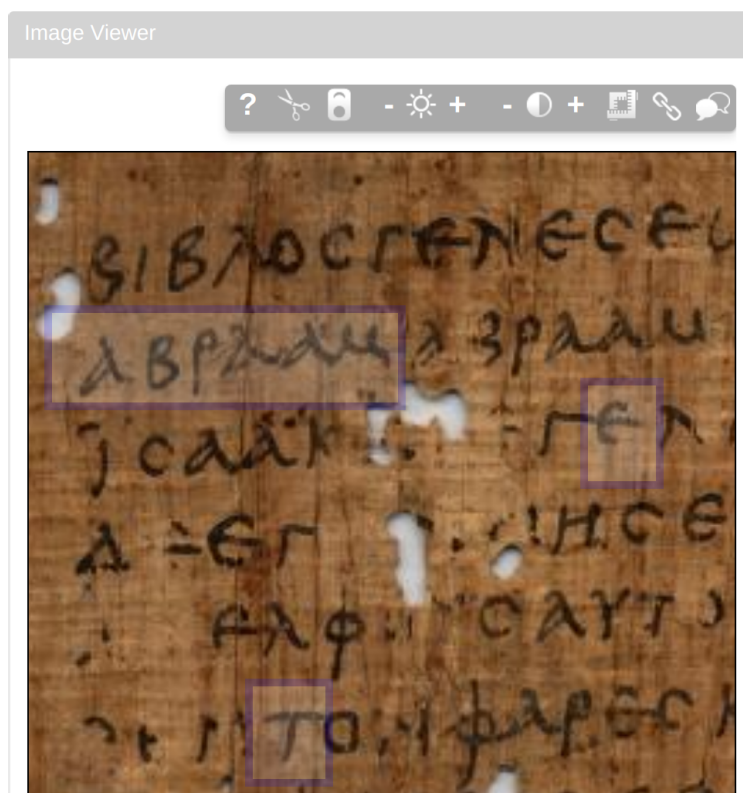


Image Viewer Gadget

The toolbar at the top of the viewer, consists of these controls, from left to right:

Help - Shows basic help information about how to navigate an image within the viewer.

Clip - Extracts a selected region box from the current image and displays a popup with a new image of the clipped region which can be copied or saved for external use. Some institutes have asked for the INTF to disallow more than 5% of the image to be clipped and downloaded. This protection is turned on for an image by raising the

¹⁹⁹ International Image Interoperability Framework™; IIIF Consortium, “IIIF | International Image Interoperability Framework,” IIIF, 2018, <http://iiif.io>.

‘PROTECTLEVEL’ for an image entry in the SURROGATEIMAGE table.

Invert - Inverts the pixel data for the image, giving another perspective of the image which may help identify text on a page. Sometime also, negative film of manuscripts has been digitized and the invert function can give a positive view of a digitized negative.

Brightness - The next three controls allow adjusting of the brightness of the image.

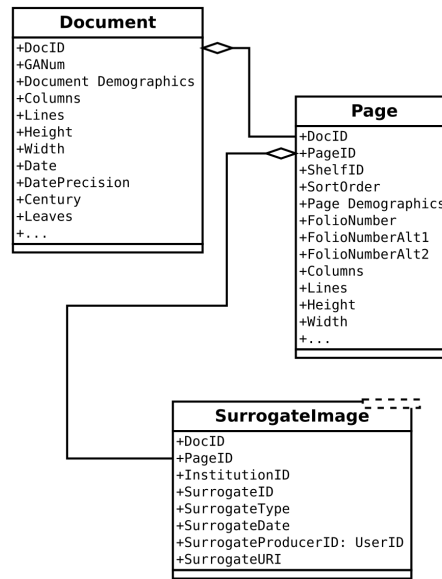
Contrast - Three contrast controls follow the brightness adjustments.

Region Box - The region box function allows the user to annotate a manuscript image by highlighting a box around a point of interest. Multiple boxes can be drawn and adjusted.

Persistent Link - Continuing a core principle within the software system, the external link button will produce a persistent URL which can be used to refer another reader to the current view of the manuscript image, including zoom level, pan location, and any region box annotations added to the image.

Discuss Button - Also a consistent theme within the software system, the discuss button will create an appropriate draft message within the proper context of the message forum for discussing the current image. The draft message includes the URL produced from the Persistent Link control, preserving the view for every message reader, and files the message under sub-categories for the appropriate manuscript and manuscript page, keeping together all discussion about similar context.

4.3.4. Page and Image data model



Page and Image, Data Model Excerpt

Just as Documents have a one-to-many relationship with Pages, so also Pages have a one-to-many relationship with SurrogateImages. In the excerpt of data below, we can see 2 pages of manuscript 10046, having 2 images associated with each page:

SurrogateImage Table Excerpt

DocID	PageID	SurrogateURI
10046	130	http://ntmss.info/images/webfriendly/10046/10046_013_INTF.jpg
10046	130	http://ntmss.info/images/webfriendly/10046/10046x6238_30JP2_UMich.jpg
10046	140	http://ntmss.info/images/webfriendly/10046/10046_014_INTF.jpg

10046	140	http://ntmss.info/images/webfriendly/10046/10046x6238_31JP2_UMich.jpg
-------	-----	---

There are many reasons why multiple images might be desired for a single page. An earlier black and white photograph may have captured the artifact in a better state than a later full color photo. Different lighting might be applied during photography to highlight different inks used in the manuscript. Sometimes the fragments which make up a manuscript page reside at different holding institutions and thus separate photos may be taken of each fragment. For whatever reason, the software system supports as many images as desired for each page.

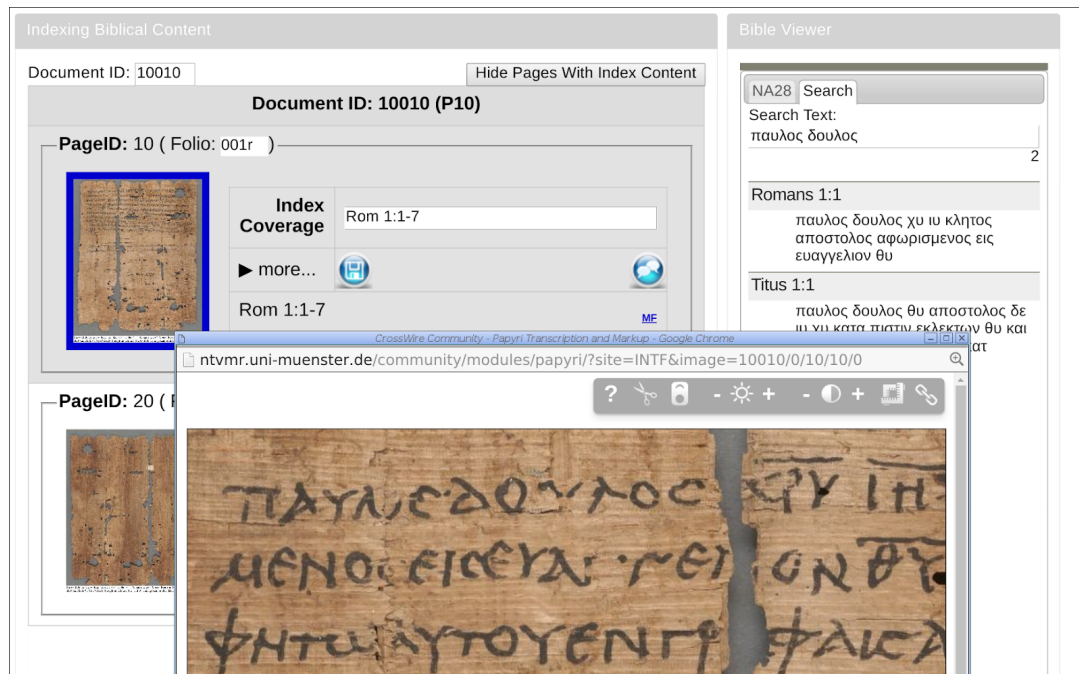
4.4. Indexing folios of manuscripts

The institutions involved in the process of editing the ECM refer to as “Indexing” their task to identify the text scope of the content for each folio side of a manuscript, i.e., recording which Bible verses are present on each page. Indexing is performed after images have been digitized ([section 4.3](#)) and before the work of transcription ([section 4.5](#)). Indexing can be done by individuals who have not yet attained the necessary paleography skills to transcribe; it serves as a natural stage in the training process.

Indexing produces the data required to seed the transcription process with the correct portion of base text material for each folio side, making the transcription process more like the traditional collation process where a researcher only is required to identify a manuscript’s differences to a base text. Since indexing can be done much quicker than a full transcription, for the ECM project indexing is often done years before the same

material will be transcribed. During this time before transcription is performed, the index data enables pages to still be included in search results for a biblical scope, even though the folio has not yet been transcribed.

4.4.1. The Indexing gadget



Indexing Workspace

As seen from the workspace page shown above, the Indexing gadget in the VMR CRE is usefully paired with the search features in the Bible Viewer gadget to assist the researcher in finding the content they discover on a manuscript page. Clicking on the thumbnail of a page will open the full image of that page for viewing. Typing a few words seen on the manuscript page into the “Search Text” field of the Bible Viewer gadget can often narrow the possible options down to a manageable set of points in the

text. Once the beginning and ending verses are identified, the researcher enters the range or list of ranges into the “Index Coverage” field of the Indexing gadget, recording the indexing data for that manuscript page. Along with indexing the pages of a manuscript, while a researcher is progressing through the pages of a document, they often discover features of a manuscript page which they would like to tag.²⁰⁰ The most common features are accessible in the “► more...” expansion for each page, shown expanded (now labeled “less”) below.

PageID: 10 (Folio: 001r)

Index Coverage Rom 1:1-7

▼ less...

Illu HePi CaTa KeLi Hyp Colo Lecldx Blank Noldx ComTx Help

Notes

Rom 1:1-7 ME

Indexing a Manuscript Page, Quick Feature Tags Shown

The quick-feature tags consist of the following, from left to right:


- Illuminations
- Headpieces
- Canon Tables
- Kephalaia List


²⁰⁰ See section [2.5. Tagging Features](#)

- Hypothesis
- Colophon
- Lection Index
- Blank Manuscript Page
- No Indexed Content
- Commentary Text
- Help, Indexing Issue, please review


This quick-feature tool encourages the capture of additional data during indexing work by offering the most common tags without disrupting the workflow of the indexer. A feature tag can be quickly added with click of a checkbox. A few of these feature tags assist in management of the indexing task. “No Indexed Content” will mark the page as complete with regard to the progress of indexing the manuscript. Without this feature, there would be no way to differentiate between whether a page has no indexable content or simply has not yet been indexed; the work assignment would never get to 100%. The “Help” feature allows a researcher to mark a page as needing assistance. An indexing manager can then search for all pages with this feature tag and provide the necessary help. The most common and time consuming scenario is identifying biblical content on a page filled with commentary text. If a commentary page has been thoroughly reviewed and determined to have no biblical content, though it has text on the page, the “Commentary Text” feature can additionally alert a future researcher that a thorough examination of the page has been completed and no biblical text has been found.

Besides the quick-feature tagging facility, notice also that the folio number for the page can be added or adjusted, notes from the indexer can be included, and the identity of the researcher who added the index data is captured in the lower right corner.


In tandem with the “Help” feature, the discuss () button can be used to initiate a message on the Indexing Discussion forum where others might offer comments on the page in scope-- possibly a discussion whether a commentary manuscript or lectionary is citing Mark or Luke, or a comment from a reader who notices that indexing data for a page is incomplete or incorrect.



WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER



INSTITUT FÜR
NEUTESTAMENTLICHE
TEXTFORSCHUNG




[Welcome](#)
[Manuscript Workspace](#)
[NT Transcripts](#)
[NT Conjectures](#)
[Forum](#)
[Liste](#)
[Transcribing](#)
[Indexing](#)
[Status](#)
[Help](#)
[NA28](#)
[About](#)
[Collation](#)

[Message Boards Home](#) [Recent Posts](#) [My Posts](#) [My Subscriptions](#) [Statistics](#) [Banned Users](#) [Search](#)

[←](#) **32110 (2110) Page 8550 (folio 423r) Indexing**

[Threads](#) [Previous](#) [Next](#) [Post New Thread](#) [Permissions](#) [RSS](#) [Subscribe](#) [Lock Thread](#) [Move Thread](#) [Move to the Recycle Bin](#)

[Matthew Solomon](#)
32110 (2110) Page 8550 (folio 423r) Indexing
8/4/17 9:04 PM
[Reply](#) [Reply with Quote](#) [Quick Reply](#)

[Jump To Artifact](#)
This folio also includes Heb 13:23-25 as well as 1Tim 1:1-2. Can someone update this?

0 (0 Votes) [Flag](#) [Edit](#) [Top](#)

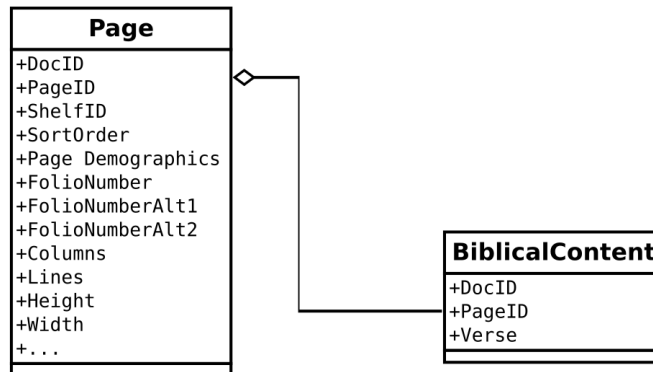
Powered By [Liferay](#)

[Settings](#) [Online Friends \(4\)](#)

Discuss Post, Indexing Forum

Notice the pre-populated “[Jump To Artifact](#)” link which returns the forum reader of the message to the exact manuscript and page context in the indexing facility.

4.4.2. Indexing data model



Indexing, Data Model Excerpt

The relevant portion of the VMR CRE data model involves the Page and BiblicalContent tables. A manuscript page can have none or many verses associated with it and a verse can have none or many manuscript pages associated with it. Consider the BiblicalContent table sample shown below.

BiblicalContent Table Sample

DocID	PageID	Verse
10010	10	Rom.1.1
10010	10	Rom.1.2
10010	10	Rom.1.3

10026	10	Rom.1.3
20001	1230	Rom.1.3

We can see that if we query:

```
SELECT Verse FROM BiblicalContent WHERE DocID=10010 AND PageID=10;
```

we will discover that Page 10 of manuscript 10010 contains Romans 1:1-3.

If we query:

```
SELECT DocID, PageID FROM BiblicalContent WHERE Verse='Rom.1.3';
```

we will discover that Romans 1:3 can be found on 3 manuscript pages.

4.4.3. History of indexing tools in the VMR CRE

In 2009, the Center for the Study of New Testament Manuscripts (CSNTM) (<http://csntm.org>) was beginning to amass a valuable collection of manuscript images; however, their project was lacking the necessary index information about these images to make their collection practically useful. Without index information to provide what biblical content was contained on each image, a user was forced to wade through sometime hundreds of raw jpeg files to find the specific folio of interest. A partnership between CSNTM and CrossWire was formed to apply CrossWire's community framework tools, which were the very beginnings of this present research project, to a new online project allowing scholar volunteers the opportunity to claim responsibility for and record the index data about each folio image for a manuscript. For this work, a new component for the community framework was developed for folio indexing. This

first incarnation of the component interfaced directly with the CSNTM site to discover which manuscripts and images where available, would generate thumbnails on the fly and build a dynamic status page of work needing to be completed. The status page communicated if any user had claimed responsibility for each manuscript, and if so, progress bars would convey the percentage of completion. If a manuscript was unassigned, a button would be offered to let a willing volunteer sign up to do the work. This was the next stage in transition from the status page developed for the KJV2003 effort (see [section 2.1.1](#)) to a more general use project status utility, shown below, which portrays the infant stages of what is now the Project Status gadget (see [section 4.0.2](#)).

Status	Indexer	Gregory-Aland Number	Type	Image Count	CSNTM Info
	Claim Responsibility	GA 0309	Majuscule	2	View Info
40% Complete (View)	shedii	GA 0311	Majuscule	10	View Info
100% Complete (View)	labarum49	GA 0312	Majuscule	14	View Info
0% Complete (View)	jtsmelsner	GA 0313	Majuscule	7	View Info
100% Complete (View)	wdilbeck	GA 0314	Majuscule	10	View Info
0% Complete (View)	gmhagues	GA 0315	Majuscule	22	View Info
	Claim Responsibility	GA 0322	Majuscule	4	View Info
	Claim Responsibility	GA 6	Minuscule	510	View Info
	Claim Responsibility	GA 8	Minuscule	0	View Info
	Claim Responsibility	GA 9	Minuscule	0	View Info
	Claim Responsibility	GA 11	Minuscule	0	View Info
	Claim Responsibility	GA 28	Minuscule	306	View Info
	Claim Responsibility	GA 43	Minuscule	0	View Info
	Claim Responsibility	GA 60	Minuscule	1	View Info
	Claim Responsibility	GA 65	Minuscule	0	View Info
100% Complete (View)	ansak	GA 69	Minuscule	426	View Info
	Claim Responsibility	GA 72	Minuscule	0	View Info

Indexing Status Tool, Infant Stages

Users could see a list of their assigned work and by clicking on one of their assignments they would be presented with a table of thumbnails representing each jpeg available for that manuscript. Next to each thumbnail was a data entry box allowing

the input of index data for each folio.



The screenshot displays a web-based indexing tool with three identical forms stacked vertically. Each form is titled with a file name: 'GA_043_0010b.jpg', 'GA_043_0011a.jpg', and 'GA_043_0011b.jpg'. Each form contains a small thumbnail image of a manuscript page on the left. To the right of the thumbnail are two text input fields labeled 'Description' and 'Biblical Coverage', followed by a 'Save' button. The forms are separated by thin horizontal lines.

Indexing Tool, Infant Stages

When the thumbnail was clicked, the manuscript page image viewer would be presented in a new window to show the full size image for that folio. CSNTM desired to have the index information available in PDF format, the status page also included a link to download a PDF of the supplied index information, which was generated dynamically from the data contributed by the assigned scholar. The collaboration saw 60 scholars volunteer to contribute index data for 3989 folios.

The CSNTM partnership was not without issue. CSNTM would occasionally switch URLs for their manuscripts. Establishing a mechanism to deal with these changes was important to keep the two projects in sync. When the indexing tools and work migrated to the NTVMR at the INTF in Münster, CSNTM voiced approval, but later, the obstacle of image permission arose. The CSNTM had obtained permission from

holding institutions to make images available. When the tools were hosted at CrossWire, deep linking to display the images from the CrossWire domain was seen as technical assistance to the CSNTM; however, the same deep linking of images into the system at the INFT web domain posed questions about whether any violation might be perceived by the holding institutions. To avoid any objections from those who had granted CSNTM image display permission, most image permissions from CSNTM were revoked. This caused the INTF to begin replacing the new hi-resolution images from CSNTM with digitizations of the same manuscript from the INTF's black and white microfiche collection.

4.5. Transcribing

After witnesses are catalogued, images gathered and indexed, the work to produce a critical edition moves on to the process of creating for each chosen witness, a

“diplomatic transcription [which] is one step removed from the typographic facsimile. The editor uses carefully chosen critical symbols or abbreviations to indicate details of inscription such as interlineations and cancellations instead of reproducing their physical appearance in the original. Editors of diplomatic transcriptions often standardize the placement of such routine elements of the source text as datelines, greetings, salutations, titles, and the indentation of paragraphs, and they may also supply missing punctuation, expand ambiguous or archaic abbreviations and contractions, or even supply words unintentionally omitted by the author or destroyed by mutilation of the original source text. None of these corrections or emendations, however, is made silently: each is given within a form of brackets indicating such editorial activities.”²⁰¹

Transcription work in the VMR CRE is done using a gadget built on the work of Martin Sievers’ team, including Yu Gan, at the University of Trier, as their contribution²⁰² to the Workspace for Collaborative Editing, in collaboration with this project and ITSEE in Birmingham. This Online Transcription Editor (OTE)²⁰³ centers around writing a

²⁰¹ Mary-Jo Kline and Susan Holbrook Perdue, “Diplomatic Transcriptions,” *A Guide to Documentary Editing*, 3d edition, accessed 2017, <http://gde.upress.virginia.edu/05-gde.html#h2.4>.

²⁰² *Kompetenzzentrum Trier*, “Workspace for Collaborative Editing,” *Universität Trier*, 2018, <http://kompetenzzentrum.uni-trier.de/de/projekte/projekte/collaborative-editing/>.

²⁰³ Source code for the editor can be found on SourceForge: *Kompetenzzentrum Trier*, “Online Transcription Editor (OTE),” SourceForge, 2017, <https://sourceforge.net/projects/wfce-ote/>.

what-you-see-is-what-you-mean (WYSIWYM)²⁰⁴ web-based editor for the TEI schema, inspired by EpiDoc²⁰⁵ used for this present research project. This TEI schema development was led by ITSEE's Hugh Houghton and agreed to by the ECM editorial projects. Trier based their work on the popular TinyMCE²⁰⁶ HTML editor. The OTE includes menus and dialogs to assist the researcher with composing a transcription. The content may then be obtained from the OTE with a simple call asking for the TEI markup. It has been an important goal of this research project to provide a transcription editor to the humanities scholar which would shield them from the intricacies of TEI XML markup. The transcriber is never asked to learn TEI, yet produces highly complex and rich TEI markup naturally by using the OTE as they would a specialized word processor. The TEI markup, produced by the editing tool, is regular and conformant to the standards of the project, regardless of the transcriber. In contrast, a project which relies on the transcribing scholar to tediously hand edit markup in conformance to a project's stated guidelines for using the TEI, inevitably results in irregular data output from different transcribers and certainly incompatible usage between projects with differing guidelines. With TEI markup generated uniformly by the OTE, the potential exists for multiple projects to integrate the same transcription editing software to produce perfectly compatible TEI markup. For more on this potential advantage, see [section 3.4.2. TEI](#).

²⁰⁴ For more on WYSIWYM, see [section 3.4.2.1](#).

²⁰⁵ Tom Elliott, Gabriel Bodard, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, Scott Vanderbilt, et al. (2007-2017), "EpiDoc Guidelines: Ancient documents in TEI XML (Version 8)," The Stoa Consortium, 2017, <http://www.stoa.org/epidoc/gl/latest/>.

²⁰⁶ Ephox, "TinyMCE - Full featured web editing. Featherweight download," TinyMCE, 2018, <http://tinymce.org>.

4.5.1. Transcription gadget

The VMR CRE wraps the OTE into a gadget which saves TEI content, returned from the OTE, into a versioned transcription repository.²⁰⁷ A user may have access to create and edit their own personal transcriptions (= initial), project-wide transcriptions (= reconciled), or site-wide transcriptions (= published)-- each having version history. The VMR CRE Transcription gadget also includes a palaeography tool to assist a transcriber when encountering rare symbols, abbreviation, or ligatures. If a portion of the unknown text cannot be easily identified, the researcher can enter one or more letters into the palaeography assistance tool and will be presented with images of text instances elsewhere which including these letters, offering possibilities. As more and more rare text graphemes are tagged, the system grows more helpful.

The screenshot shows the 'Transcription Workspace' interface. The main area contains a transcription of Greek text with various tags and a 'Paleography' section below it. A sidebar on the right displays a table of document pages.

Transcription:

PB 1 Note

CB

1 1 βιβλος γένεσε (ως) τω χυ ου δαυιδ

αβρααμ 2 αβραμ ξγ ξγνησε ν το ν [ισαακ]

ισαακ δ [ε] ξ γεν νη σεν τ [ον] το [αβ]

δε εγ [ε] νν ησε ν τ ο ν το υ δαν κ [α] ι τ [ους]

α [δ] ελφ ου ς αυτου 3 ιουδ ας δ ε γε ννη

σεν τον φαρεσ και τον ζαρε εκ της θ α

Paleography

Images of various graphemes: * κ *, και, εικοσι, εκκλησιας, γλυκυ, κλειδε, κλεις, εσκοτομενη, κυκλω, σπηνεγκ.

Version History: κ Publish Preview Last Saved From Basetext

Manuscripts **Pages** **Favorites**

Doc ID: 10001

Page Id	Content	Image
10 (1r)	Matt inscriptio-1:11	
20 (1v)	Matt 1:14-23	

Page Id Content Image

Transcription Workspace

²⁰⁷ For more on versioned repositories, see [section 3.4.3](#).

4.5.2. Transcription reconciliation

Quality assurance practices for the ECM require that a transcription for a manuscript be produced independently by two transcribers.²⁰⁸ The products are then compared to each other where differences are reviewed by a manager and reconciled to produce a final transcription. Previously, ITSEE and the INTF used a comparison tool called Collate, written by Peter Robinson for the reconciliation process.²⁰⁹

Transcription reconciliation is an important phase in the workflow to produce a critical edition. It forms the QA process for transcribing and is a discrete step performed by an independent agent other than the initial transcribing persons. In this step, the reconciler takes two independently transcribed efforts of the same artifact and compares them to each other. For each point of difference between the initial transcriptions, the reconciler will consult the original manuscript and choose the correct rendition from one of the two initial transcriptions. Once all points of difference have been reviewed and reconciled, a third transcription is stored, consisting primarily of the points of agreement in the two initial transcriptions, along with the reconciled points of difference. This third transcription is called the reconciled or project transcription.

4.5.2.1. Reconciliation gadget

²⁰⁸ David C. Parker, *An Introduction to the New Testament Manuscripts and Their Texts*, 1 edition (Cambridge, UK; New York: Cambridge University Press, 2008). 217.

²⁰⁹ Ibid. Also, for more on the Collate program by Peter Robinson, see [section 4.6.2.1](#).

The VMR CRE Reconciliation gadget has been built starting with the work from an open-source project named Mergely.²¹⁰ Mergely provides web-based tools for displaying and reconciling differences between two electronic buffers of text.²¹¹ The Reconciliation gadget is project-aware, knowing who is assigned to a manuscript based on from which project page the Reconciliation gadget is being used. The gadget subscribes to page select events so it can be used with the Catalogue gadget, and will respond by loading into each window for comparison the transcriptions for that manuscript page done by the assigned project members. The master transcription is shown on the left and a reconciliation tool is offered to move changes from right to left. Live editing is also allowed in the left window. The gadget will publish an event to notify when the master transcription has been updated. Coupling on the same page the Reconciliation gadget with the Transcription Editor gadget, which will listen for (= subscribes to)²¹² this update event and show the WYSIWYM²¹³ display of the master transcription, a nice reconciliation workspace can be easily constructed with the following three gadget layout.

²¹⁰ Jamie Peabody, “Diff text documents online with Mergely, an editor and HTML5 javascript library,” Mergely, 2018, <http://www.mergely.com>.

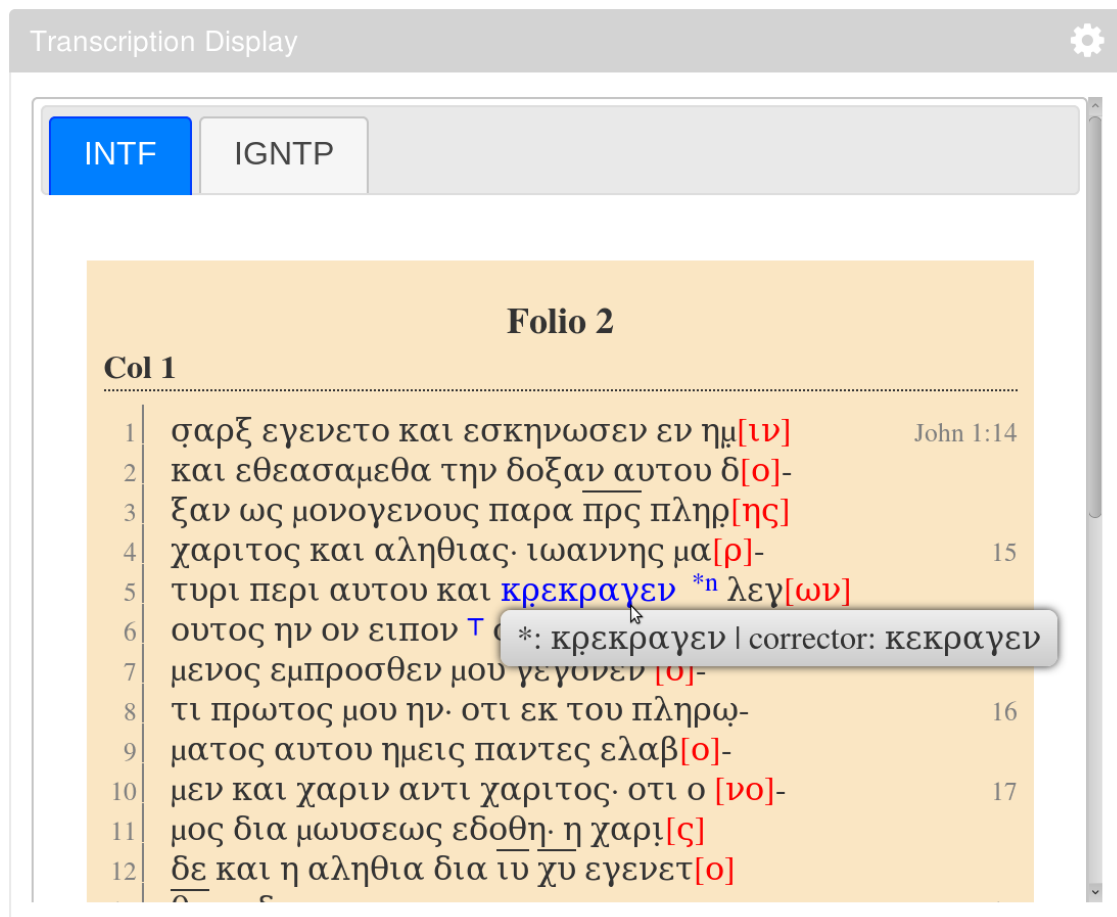
²¹¹ Mergely can work with against any two memory buffers, and for the reconciliation process in the VMR CRE these buffers are retrieved from the transcription/get web service, querying the service for a window of data for one manuscript page.

²¹² For more on the publish/subscribe architecture, see [section 3.5.2. Inter-gadget communication](#).

²¹³ For more on WYSIWYM, see [section 3.4.2.1](#).

4.5.3. Transcription viewer gadget

The transcription viewer gadget is a simple display gadget which allows transcription viewing of any selected manuscript page.



Transcription Display Gadget

The mouse can be hovered over notes, corrector readings, and other elements to display detail information. Unsure letters are marked with an underdot. Supplied segments of text will be seen bracketed in red. Lines which end in a partial word which continues on the next line will have a trailing '-' symbol. Along the right margin, standing off

from the transcription, biblical verse references help orient the reader. Tabs at the top show when multiple transcriptions are available. The gadget can be configured to always show a particular user's personal transcription, which can be useful when a reader wishes to setup a public page to publish research from their own work.

4.6. Regularizing and Collating

Collating manuscripts refers to the comparison of multiple witnesses at a single point in a text and the visualization of those differences. Manuscript witnesses to a text often have many quirks which can be contributed to the writing style of the scribe but not necessarily relevant to a critical edition of a work. Depending on how detailed a project wishes to transcribe with respect to orthography, there may be a need to provide a “regular” alternative form to these unimportant witness deviations in a transcription. This act is called “regularizing” and this stage of work in producing a critical edition is called “regularization”.

4.6.1. Regularization

This process is necessary to allow a project to specify which variants are deemed unimportant for the reader of a critical edition and thus will not be noted in an apparatus. Since these differences are obvious during the collation phase of a project, regularization and collation are done hand in hand. Many types of differences which a project may decide not to include in a critical edition are common and have proper names to describe the entity. Some Greek examples include:

- Moveable nu - These sometimes occur when the last word ending a line also ends in the Greek ν (nu) consonant and is written simply as a supralinear stroke above and often extending just past the penultimate character in the word. If

this is captured in the transcription, it would be regularized as simply the greek letter ν (nu).



- καὶ compendium - This symbol is shorthand for the very common Greek word καὶ and would be regularized as such.



- Abbreviation - Often words are abbreviated in a manuscript. These are regularized to their expanded full word.
- *Nomen sacrum* - Many names in the Greek New Testament, such as the name of God, are given honor by shortening the name and highlighting with a supralinear stroke. Below is an example of this for the name of Jesus in the Greek.



- Iotacism - As pronunciation of Greek language changed over the course of time, words also changed in spelling to more closely match the change in pronunciation. Diphthongs including an iota were often shortened to simply the iota. These spelling differences, if unambiguous, are regularized to the spelling which will be used in the critical edition's baseline text.

- Spelling difference - Often the spellings of words by scribes in a textual tradition vary, sometimes significantly in the case of proper names, and if there is no question to the intent of the scribe, these spelling variations are typically not annotated in the critical edition.

When a regularized form is included for a word in a manuscript witness, the collation process will use the regularized form instead of the orthographic form and thus not report differences for these types of variation. Often global regularization rules can be written and applied to an entire manuscript repository. This is true for *nomen sacrum* in the ECM project. It may be that the irregular form of a word in a text, which a regularization rule will be constructed to regularize, can be a regular form elsewhere. In these instances, a regularization will be localized to a single reading in a specific manuscript. Without regularization, a collation of hundreds of witnesses to a single point in a text would display more variation noise than variation of interest. To see this exemplified, compare the two Variant Graph displays presented in [section 3.9.5](#).

4.6.2. Collation and its history

Before fully digitized images of manuscripts and computer assisted witness comparison was available, manuscripts were compared by hand to a base text and all discovered deviations from the base text were compiled into what was then called a manuscript collation.²¹⁴ This practice was the standard way of recording the text of a manuscript

²¹⁴ David C. Parker, *An Introduction to the New Testament Manuscripts and Their Texts*, 1 edition (Cambridge, UK; New York: Cambridge University Press, 2008), 95ff.

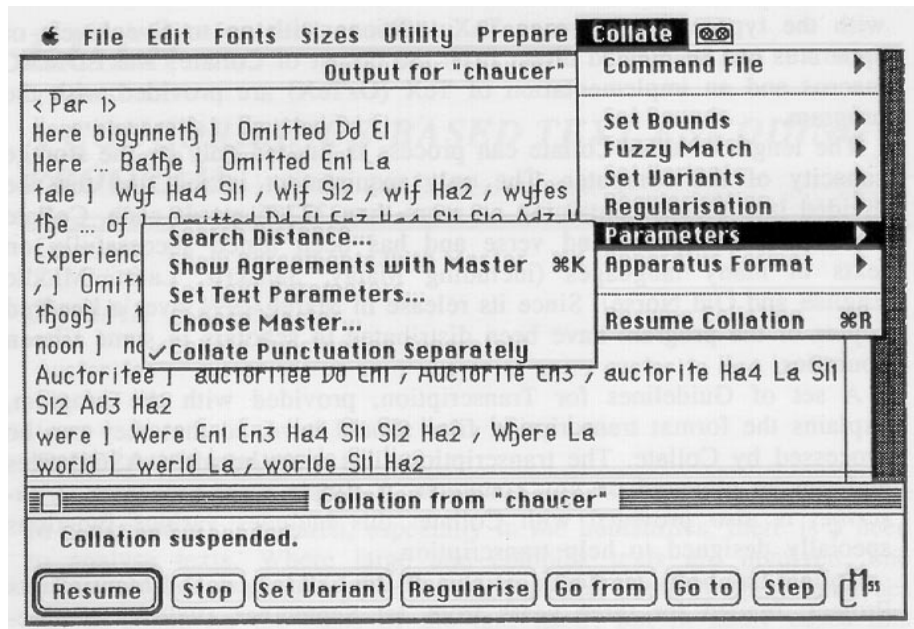
before the adoption of electronic tools and still is often the practice when viewing time with an artifact is limited. It is important to not confuse this practice-- now often referred to as a “paper collation”, with the more common usage of the term “collation” in the digital age: collation now more commonly refers to the computerized comparison of multiple electronic witnesses to the same point in a text, resulting in a display showing the differences.

4.6.2.1. Collate

The tool used at the INTF for computerized manuscript transcription comparison for the previous decade²¹⁵ to the start of this present research was written by Peter Robinson while he was serving as Research Officer for the Computers and Manuscripts Project within the Oxford University Computing Service and as chair of the Textual Criticism working group of the Text Encoding Initiative. Robinson adopted the term “collation” and used it as his program’s namesake: Collate. His work was the result of his experience collating forty-six manuscripts of the two neo-Eddic Old Norse poems *Gróugaldr* and *Fjölsvinnsmál*.²¹⁶

²¹⁵ A fun blog post on the history of collation devices can be found by Wesley Wraabe, “Collation in Scholarly Editing: An Introduction,” *Fill His Head First with a Thousand Questions*, July 26, 2008, <https://wraabe.wordpress.com/2008/07/26/collation-in-scholarly-editing-an-introduction-draft/>.

²¹⁶ Peter M. W. Robinson, “The Collation and Textual Criticism of Icelandic Manuscripts (I): Collation,” *Literary and Linguistic Computing* 4, no. 2 (April 1, 1989): 99–105, doi:10.1093/lc/4.2.99.



Collate, by Peter Robinson

For more on the history of Robinson's Collate software, see his blog post at:

<http://www.sd-editions.com/blog/?p=15>

4.6.2.2. CollateX and Interedition.

In 2008, the desire to replace the aging and Macintosh-only Collate software with a modern and cross-platform collation engine component was identified as the first common need among all participants of the Interedition effort (see [section 1.5.2.3](#)).

While Collate worked well to compare many witnesses to any chosen base text, a more general research tool was targeted to perform 'baseless' collation of any number of witnesses to each other. There is a significant and fundamental difference between a baseless collation, which graphs the variants among a group of witnesses, comparing all witnesses to each other, and what was performed previously by Collate-- performing

n-1 collations, which each compare Witness₁ : Witness_[2..n]. The latter is relatively straightforward. The former is a theoretical and technical challenge worth the distinction. For the scholar, baseless collation brings the benefit that no single text must be chosen before collation can be performed. The distinction between base and baseless collation is not merely the display of the end result; though, a baseless collation can also be displayed in the same way as variation against a single witness, if desired. The CollateX engine²¹⁷ is the realization of the Interedition effort to build this cross-platform replacement for Collate. Ronald Dekker and Gregor Middell played significant roles in its inception, guided by input from Peter Robinson. CollateX has the following component model and execution flow:

- Tokenize – Separate a witness into tokens at the desired level of granularity (word, morpheme, glyph, etc.).
- Normalize – Snap tokens to their normalized forms (expand abbreviations / contractions / *nomina sacra*, remove punctuation, etc.); i.e., regularize.
- Align – Build a baseless difference graph from all witnesses.
- Analyse – Determine corrections, omissions, transpositions, et al.

The engine provides a default implementation for each of these components, and also allows individual projects to replace each component with a domain-specific implementation. CollateX simply covers the logic of collating text; other features of a manuscript, such as page layout, which might explain a scribal variation, lie outside the

²¹⁷ Interedition Development Group, “CollateX – Software for Collating Textual Sources,” CollateX, 2018, <https://collatex.net>.

domain of CollateX.

This present project first created a simple interactive proof of concept which allowed the invocation of the CollateX engine with up to 4 distinct witnesses, repeated a variable number of times. The output is a very simple display of the results from the engine as either an alignment table or an apparatus of the analysis. The proof of concept uses all of the default components of the engine— none has yet been replaced to facilitate the unique domain of New Testament textual criticism. The proof of concept assured that the CollateX engine was mature enough for this project to use as its collation engine. The proof of concept can still be exercised here:

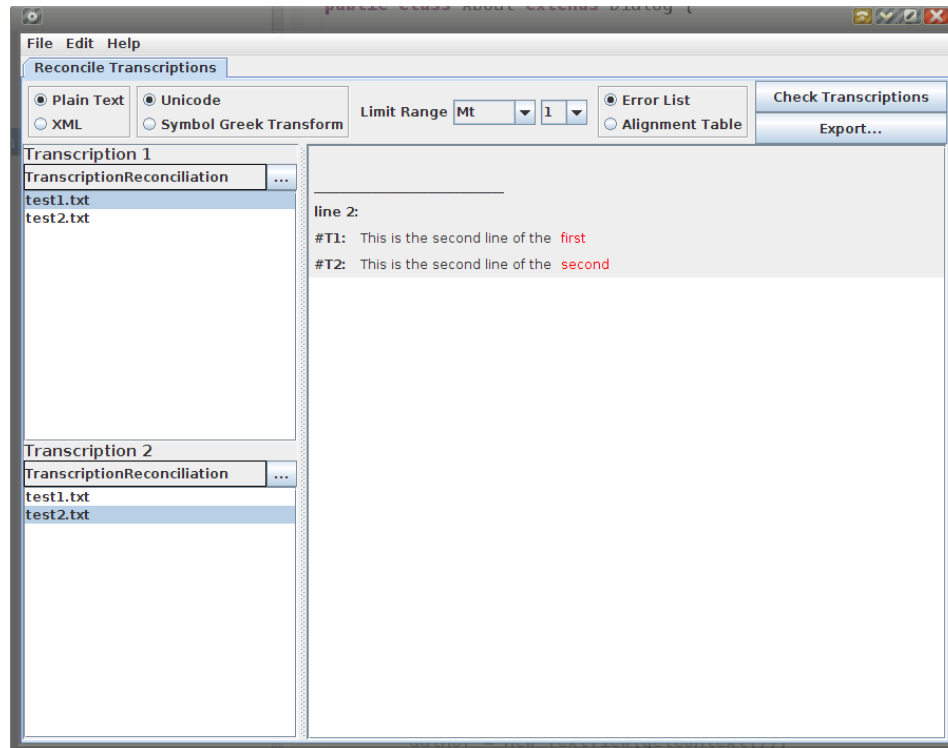
program: http://crosswire.org/~scribe/collate_demo.jsp

code: http://crosswire.org/~scribe/collate_demo.jsp.txt

The collaboration with the Interedition effort resulted in a modern, cross-platform replacement to Collate for the computerized comparison of texts. This effort, with Peter Robinson's consultation has now become the successful CollateX project, the fruits of which are used as the collation engine by many research efforts.

4.6.2.3. CollateX for transcription reconciliation

In the previous workflow at the INTF, Collate was also used as the tool to perform reconciliation between two initial transcriptions (see [section 4.5.2. Transcription reconciliation](#)). This history prompted a second proof of concept, a transcription reconciliation tool which used the CollateX engine, to test whether it could be used as a replacement for the reconciliation task at both institutes.



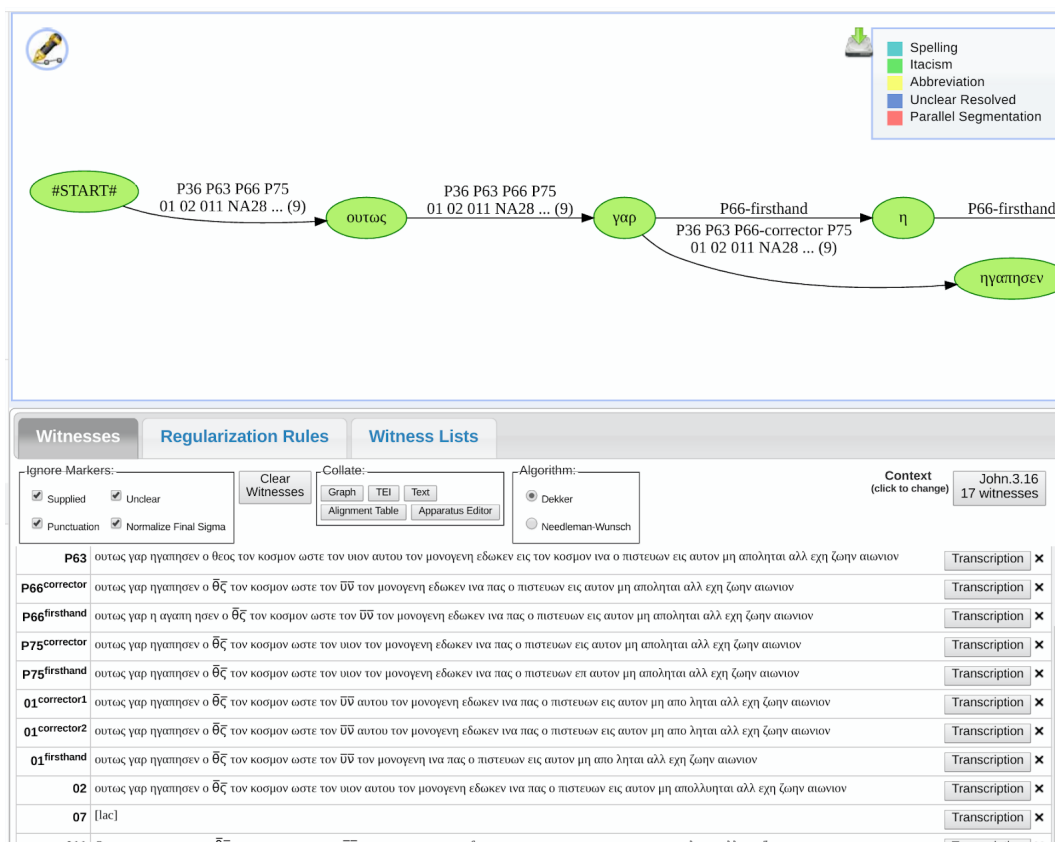
Reconciliation Proof of Concept

This path proposed by the proof of concept was never pursued, in favor of the web-based Mergely solution described in [section 4.5.2.1](#).

While CollateX successfully produced the workhorse behind text comparison, Robinson's Collate software provided much more than just a collation engine; it included user-facing tools for regularization and other work involved in producing a critical edition. The following section describes the user interface initially developed for regularization within the VMR CRE.

4.6.3. Collation and Regularization gadgets

As collation and regularization are done in practice together as a single work process, the VMR implements both features as a single tool with three use cases. The first use case involves the functionality for choosing a set of witnesses along with a desired text range, resulting in a visualization of the differences across the specified parameters.




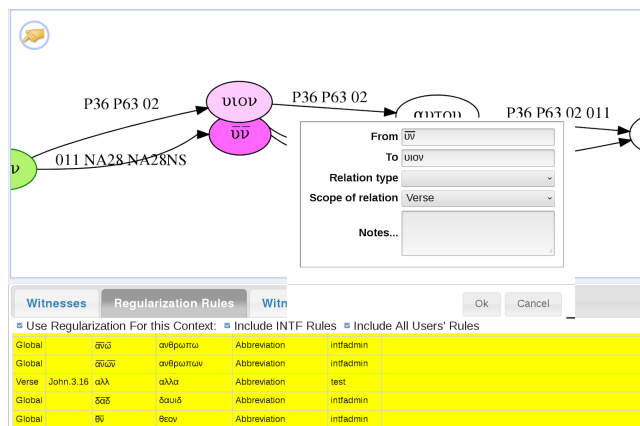
Regularization Workspace

The second use case is centered on building regularization rules for manuscripts. As a scholar seeks to collate many witnesses for a text, she will inevitably find variants from the collation output which are not interesting for her research-- often spelling and

orthographic anomalies in a witness. These differences need to be factored out of the collation result by adding to a regularization ruleset. This iterative process involves:

- regularizing witnesses against the ruleset,
- collating the regularized witnesses,
- displaying the collation,
- allowing the scholar to select variants which are not interesting and should be factored out of the collation,
- adding a new rule to the regularization ruleset which represents the scholar's decision,
- *repeat*.

Rules are added by first clicking the collation edit button  seen in the image above, choosing an unwanted variation in the graph, and then dragging this irregular token bubble and dropping it on top of its corresponding regular form. This will initiate the regularization rule creation dialog, allowing this form to be associated with its regular form during the next collation iteration.



Global	Verse	Text	Abbreviation	User
Global	860	avpuma	Abbreviation	infadmin
Global	860	avpuma	Abbreviation	infadmin
Verse	John.3.16	avp	Abbreviation	test
Global	865	avp	Abbreviation	infadmin
Global	86	avp	Abbreviation	infadmin

Regularization Rule Creation

The gadget allows any user-created manuscript list, and provides for all extant witnesses to be the starting point for witness selection, the list can be further narrowed by clicking the [X] to the right of any witness which should be excluded from the collation. The Interedition CollateX tool is used to produce the collation and Graphviz²¹⁸ is then invoked to generation the visualization. The visualization viewer and editor were written in collaboration with Tara Andrews from Leiden and Joris van Zundert from the Huygens Institute in the Netherlands during an Interedition focus group prototyping user interfaces for CollateX.

Once this regularization process produces from the raw manuscript transcriptions a collation showing only variation desired to be included in the edition, the stage of apparatus editing then becomes the focus of the editing process, discussed in the next section.

4.7. Editing an Apparatus

After the collation / regularization cycle is repeated to the point that only desired variants remain in the collation, the next phase of production involves editing a critical apparatus against a chosen base text. First an initial baseline text must be chosen against which an apparatus will be build. This is often done by selecting the majority reading at each variant passage, which would produce an eclectic (= constructed from the readings across various manuscripts, instead of one document) text, or simply by

²¹⁸ Graphviz, “Graphviz - Graph Visualization Software,” Graphviz, 2017, <http://www.graphviz.org>.

selecting the entire text of the best surviving witness. It should be noted that, to serve its purpose as the baseline for an apparatus, the base text must be extant at every verse in the edition. If this initial baseline text is also intended by the editors to represent their best assessment for an initial text,²¹⁹ [section 4.8](#) will offer tools for this evaluation. The ECM work for a fascicle begins with the text published in the most recent edition of the Nestle-Aland Greek New Testament.

Editing an apparatus, or as ITSEE refers to it, “editing a collation,” is done in the VMR CRE using the Collation and Apparatus Editor developed in Birmingham at the Institute for Textual Scholarship and Electronic Editing (ITSEE). This work was undertaken initially as part of the Workspace for Collaborative Editing (WCE)²²⁰ and in practice to facilitate the work to edit the Gospel of John for the ECM. Feedback for the design of the editor came from all participating members of the WCE. Catherine J. Smith, the ITSEE Technical Officer and with an academic background in the field of biblical studies, is the principal developer with David Parker as her initial primary pilot user in his work to edit John. With this project-- including Ulrich Schmid and Bruce Morrill using the tool for regularization work-- serving as a real-time proof of concept, the tool has benefited from immediate feedback from David, other ITSEE team members, and the wider ECM editing community. Development is still ongoing at the time of this writing but the tool is already fit for purpose. From the yet to be released Guidelines from ITSEE, with the collation editor, the user “regularises noise, sets and

²¹⁹ a text which represents the earliest reading from the extant transmission history.

²²⁰ ITSEE, “The Workspace for Collaborative Editing,” University of Birmingham, 2017, <https://www.birmingham.ac.uk/research/activity/itsee/projects/workspace.aspx>.

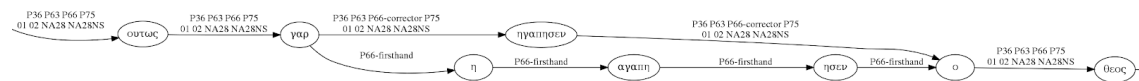
defines the variation units and places them in the correct sequence. This data is subsequently made available in other interfaces for adding versional and patristic data and is output to a database for the application of the Coherence-Based Genealogical Method.”²²¹

This tool has been incorporated into the VMR CRE in a similar way the Online Transcription Editor, built by the University of Trier, was included, though the integration was more complex. For more on this, see [section 4.7.3](#) on the Collation and Apparatus Editor gadget. Editing with ITSEE’s editor occurs across 3 phases and thus any verse can be in 1 of 4 states: regularization, set variants, order variant readings, approved. The following sections give a brief overview of these phases and states. A detailed description of each phase, covering the intricacies of this ECM editing stage within the ITSEE editor, are forthcoming in a guide compiled by Parker, Smith, and Jac Perrin, “*Editio Critica Maior*: Guidelines for Editing the CollateX Output Using the Collation Editor.”

²²¹ D.C. Parker and C.J. Smith with Jac Perrin, *Editio Critica Maior*: Guidelines For Editing The CollateX Output Using The Collation Editor; Version 1.0.

4.7.1. Setting Variant Units

While collation and regularization, in the previous editing phase, have been performed at word granularity, the first step for editing an apparatus is to now determine logical and useful units to show variation. For example, in the regularization and collation phase, variation may be represented as:



Word-level Variation

This same variation, toward a view to present more meaningful and concise variant units, might be reduced to:

Set Variants John.3.16 CET									
1	οὕτως	3	γὰρ	5	ἡγαπήσεν	7	ο	9	θῆς
	2		4		6		8		10
	▲		▲		▲		▲		▲
	a. οὕτως		a. γὰρ		a. ἡγαπήσεν		a. ο		a. θῆς
	b. null		b. om.		b. ἡ ἀγαπή ἦσεν		b. om.		b. θῆς
	c. lac				c. om.				c. om.

Setting Variants, in ITSEE's Editor

First, observe that all variation is shown against a chosen baseline, seen as the first line of Greek text and metered by a running ordinal count: even numbers successively designate each word, and odd numbers to designate the location either side of each word, allowing for the apparatus to indicate an insertion. Next, notice the grouping of word in the variant of witness P66-firsthand, which reads “ἡ ἀγαπή ἦσεν”. Here we see

this clause grouped together into a single variant unit associated to the single word in the baseline reading “ηγᾱπησεν” (position 6). Regularization can also be performed in ITSEE’s editor and in most cases, the rules entered in either tool are interoperable. This gives the editor the choice for which input tool is most ergonomic for their taste. It also allows for regularization missed during the previous phase to be conveniently entered in the Setting Variant Units phase, without changing tools. Easy drag and drop editing happens in the editor to combine and dissolve units, build overlapping units of variation and much more.

4.7.2. Re-ordering variant unit readings

Once variant units are set for a verse, the variant unit selections can be saved and the ordering of the readings found within each chosen variant unit can be ordered to that in which the scholar wishes to show in the edition. After finalizing variant unit reading order, the verse can then be reviewed and moved to final state of “approved”.

4.7.3. Collation and Apparatus Editor gadget

To facilitate the integration into the VMR CRE, of this tool developed at ITSEE, a joint effort between this present researcher and Smith was made during the summer of 2015 to abstract the communication between the editor and the Magpy²²² backend system written by Zeth Green and used in Birmingham to store transcription, regularization, and editor data for the ECM work done at ITSEE. This successful effort resulted in a single re-implementation point for efforts to adapt the editor to work in other projects.

²²² Zeth Green, “Magpy 1.0 documentation,” Magpy, 2013, <http://zeth.net/magpy/>.

Two ‘implementations’ were initially developed: one for the existing Magpy system to support the continued operation of the editor in its current environment; and second, an example implementation using “local files” which allows the editor to be used stand-alone, simply storing and retrieving data to and from local files on the user’s computer as the datastore. Although only intended as an example, this latter simple implementation is being used by other projects in Birmingham who have adopted the Collation and Apparatus Editor for use in their critical editing, and now also in Peter Robinson’s Textual Communities at the University of Saskatchewan.²²³ Now a third implementation which enables the tool to be used as a gadget in the VMR CRE has been completed. This implementation uses the VMR CRE web services API as the backend for transcriptions, regularizations, and editor data used by the tool. The API needed 2 additional facilities to handle the work:

- A versioned project data repository. This facility allows a project to save any data they would like as “key - value” pairs. Data is saved partitioned by project and separated by user. Keys can be hierarchical, e.g.,
“collation/regularised/Acts.1.12/initial”. All data is versioned and attributed to author. This enables tracing any changes made during the life of the editing process. See API section [5.1.2.4. projectmanagement/project/data](#).
- The concept of a global regularization rule exception. While both global and localized regularization rules were already supported in the VMR CRE, no allowance was made for a local exception of a global regularization rule.

²²³ Peter Robinson, “Textual Communities,” University of Saskatchewan, 2017, <http://www.textualcommunities.usask.ca>.

The VMR CRE Collation and Apparatus Editor gadget derives its context from the project site where the instance of the gadget is placed. The project, team members, managing editors, witness list are all gleaned from the same project data and managed in the same way as for a transcription project. New teams and projects are created and managed in a way familiar to existing project administrators.

4.7.4. Versional and citation evidence

It is important to note that the Collation and Apparatus Editor only deals with editing the Greek apparatus for the ECM. After this editing phase is complete, the section that follows ([4.8](#)) deals with re-evaluating the editorial baseline reading, using all of the data produced for the Greek apparatus in this stage ([4.7](#)). This is the sequential flow for editing an edition for the ECM. However, once this re-evaluation is complete, an additional tool with a similar user interface as the one described herein is used to augment the Greek apparatus by adding versional and citation entries which will appear in the finished ECM. This process of adding versional and citation evidence is not further discussed in this research, yet it should be noted that it is an important part of editing a volume of the ECM.

4.8. Re-evaluating the editorial reading

If the purpose for a project researching a critical edition is to establish the best initial reading possible from all manuscript witness evidence, as is a goal for the ECM, the final phase to complete the critical edition involves re-evaluating the chosen baseline

text which the apparatus hangs upon, in light of the statistical data which is produced after the apparatus is complete. Any change in baseline reading from this evaluation will require the apparatus at that point to be inverted to adjust for the new baseline reading.

It is important to understand the distinction between texts and manuscripts. A manuscript may be physically dated to the time or times a scribe performed the work, but to speak about the date of the text within a manuscript is a much more complex concept and often not a task one can embark upon without first explaining the problems therein. First, a manuscript may be a very recent copy of a text that is very old; for example, a modern 'scribe' might choose to produce a new manuscript using the oldest known exemplar. The result would be a recent manuscript with a relatively old text. A manuscript's original production unit might be corrected centuries later against a much older witness, making this later circulation unit a closer representation to the initial text than the original production unit of the manuscript, though it is younger.

The next anomaly which complicates the topic concerns a scribe which uses multiple exemplars while copying. A codex of the Gospels and Psalms might use one exemplar for Matthew and Luke, another for Mark, a third for John, and a separate witness for the Psalms. A second hand might correct the original scribe, checking against different masters. In addition, a great number of especially old manuscripts are lost. Hence, we cannot expect to find many direct first generation relationships between the extant

manuscript population. This introduction of multiple sources into a manuscript's final production unit and the many missing links cause 'genealogical contamination'. This is to say that a manuscript's genealogical pedigree as a whole may not be shown as a simple tree, nor often even discussed intelligibly. To overcome these difficulties, the ECM applies a common method of analysis to the witnesses of a given **reading** to determine the weight of a **reading**, as relates to its attestation to the initial text. This method was originally developed at the INTF by Gerd Mink and is called the Coherence-Based Genealogical Method (CBGM).²²⁴ Because the genealogy of the text of a manuscript cannot be analysed as a whole, due to the difficulties presented earlier, the CBGM works toward the task of building **local** stemmata at points of variation. Once a local stemma is constructed at a variation unit, the parent of the stemma tree becomes the reading at that point for the initial text. Many classical criteria are applied to determine priority readings while building a local stemma: harder readings are often favored above what would likely be a later smoother substitute, shorter readings over an expansion, divergent readings in the Gospels higher than a possible harmonization, etc., but a unique development with the CBGM includes the application of statistical comparisons of manuscripts with similar readings to shed new light on what might be a divergent reading within a closely related group.

²²⁴ Gerd Mink, "Eine Umfassende Genealogie Der Neutestamentlichen Überlieferung," *New Testament Studies* 39, no. 4 (October 1993): 481–99, doi:10.1017/S0028688500011929.

4.8.1. Components for choosing the initial reading

The INTF has partnered with the University of Köln to build electronic tool for this editorial stage. The tools are now in their final development and the intent is to integrate them within the VMR CRE in the same manner as Trier's Online Transcription Editor and ITSEE's Collation and Apparatus Editor. The Köln tools are featured in the Digital ECM in [section 6.1.2. A transparent conclusion](#). One tool related to this work and developed to assist with the selection of witnesses is described in the next section, and then after, a brief mention is made of tools used before the effort in Köln, which were used for the editing work of ECM Acts.

4.8.1.1. Pre-Genealogical Coherence statistical analysis

Analytical tools have been developed for determining a manuscript's closest ancestors based on agreements in readings. While not only useful for the CBGM, one of the CBGM Pre-Statistics tools allows a researcher to select a specific manuscript, then display the number of agreements and disagreements in the extant readings at the chosen *Teststellen* of the chosen manuscript compared with all others.


 Manuscript:

Manuscript	Agree	Agree%	Disagree	Disagree%	Extant	Extant%
1719	177	100.00%	0	0.00%	177	98.33%
1769	6	100.00%	0	0.00%	6	3.33%
P98	1	100.00%	0	0.00%	1	0.56%
1719-C	175	98.87%	2	1.13%	177	98.33%
1870	157	87.71%	20	11.17%	179	99.44%
141	157	87.71%	20	11.17%	179	99.44%
82	154	87.01%	22	12.43%	177	98.33%
1852	154	86.52%	22	12.36%	178	98.89%
919-C	134	86.45%	20	12.90%	155	86.11%
1734	154	86.03%	23	12.85%	179	99.44%
2200	134	85.90%	21	13.46%	156	86.67%
919	133	85.81%	21	13.55%	155	86.11%
699-C	154	85.56%	23	12.78%	180	100.00%
627-C	142	85.54%	21	12.65%	166	92.22%
627	141	85.45%	21	12.73%	165	91.67%
110	128	85.33%	20	13.33%	150	83.33%
1780-C	151	85.31%	25	14.12%	177	98.33%
1955	126	85.14%	21	14.19%	148	82.22%
699	153	85.00%	24	13.33%	180	100.00%
177-C	153	85.00%	24	13.33%	180	100.00%
935-C	153	85.00%	24	13.33%	180	100.00%
1728	147	84.97%	24	13.87%	173	96.11%
1755	146	84.88%	25	14.00%	173	96.11%

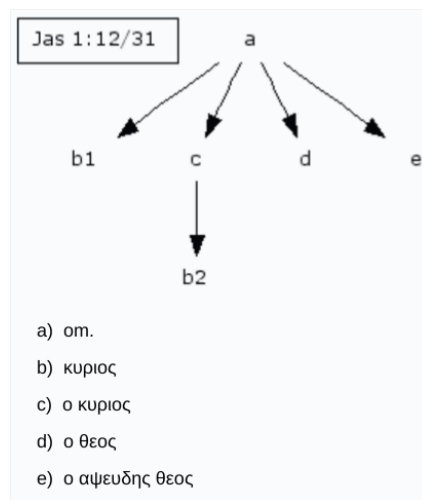
Witness Agreement Statistics Gadget

The data which feeds the algorithms used to produce the statistics computed with this component come from the editing phase for choosing witnesses. See [Section 4.2.](#)

[Choosing Witnesses.](#)

4.8.1.2. Local stemmata

At each variant unit, all classical and statistical methods are consulted, and the editor chooses a plausible local stemma for the various readings of the text at that location in the work. Once complete, the root of the local stemma determines the baseline reading to be including as the initial text reading in the edition.



Legacy CBGM Local Stemma Tool

A local stemma shows, for posterity, the weight and transmission history, as best as could be determined by the editors, for the readings behind the choice for the initial reading. It exposes much deeper than the final apparatus the minds behind and decisions leading up to the edition.

5. Open programmatic access

A web services application programming interface, or API, makes available all the functionality of the VMR CRE to users wishing to build their own unique tools and user interfaces against the system. APIs are common for software systems²²⁵ as a way to ensure connectivity and exposure for a system and its resources beyond what is developed in the system itself. In short, providing an API is simply a means for allowing others to develop unique applications which utilize the functionality-- services and data-- of an existing system. A web services API makes these facilities available over standard internet HTTP protocols which are the basis of modern Internet web browsing.

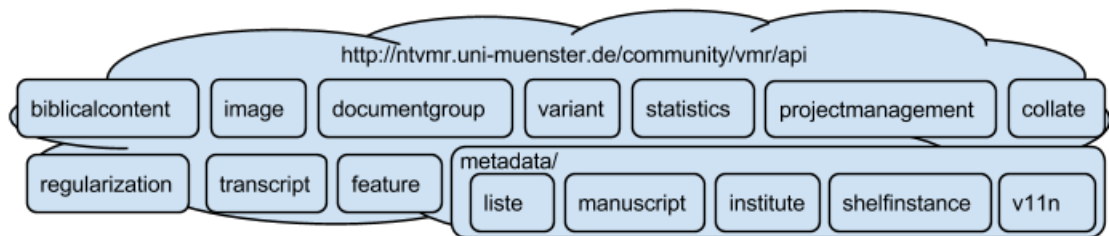
The VMR Web Services API layer is primarily useful for exposing the functionality of the VMR to other research projects wishing to access the functionality or contribute to the dataset through their own systems and tools. For more information about web services and remote invocation [see section 2.2.2](#).

²²⁵ Nearly all major software systems today include some level of application programming interface. For example, Google makes all of its APIs available here: Google, "Services," APIs Explorer, 2017, <https://developers.google.com/apis-explorer>. All major operating systems, include Windows, OSX, Linux, Android, and iOS all include APIs, without which it would be impractical for external developers to write software for these platforms.

The Web Services API for the VMR CRE instance for the New Testament Virtual Manuscript Room can be found at the following URL:

<http://ntvmr.uni-muenster.de/community/vmr/api>

Exploring the API via a web browser is possible and each API call includes a full usage page documenting its input parameters and function. A diagram of the available services can be seen in this figure:



Web Services Organization

A typical usage page for an endpoint can be seen by calling the any of the web services without pass a query string. For example:

<http://ntvmr.uni-muenster.de/community/vmr/api/image/put/>

5.1. Details of the API

The VMR CRE web services API adopts a noun/verb nomenclature with calls organized by category. This means that the last 2 segments of an API URL will first designate on which type of object the call will operate, and second of the action to be performed on the object, e.g., image/put will *put* an *image* into the VMR CRE. Any segments before the final two are merely for organizational purposes. This is different from a strict REST convention which confines the operation to one of the 9 HTTP verbs.²²⁶ The VMR CRE places no semantic meaning on the HTTP verb, as REST would do. Both GET and POST HTTP verbs are accepted as semantically identical, relegating the operation of the method call (= semantic action) instead to the final segment of the URL, just after the noun: .../noun/verb. Parameters for the request are sent to a web service call either in the query string, or as standard HTTP FORM POST parameters (application/x-www-form-urlencoded). As an example, an API request to obtain a transcription of manuscript 30093 of the Apocalypse by Darius Mueller in TEI form with no stylesheet applied would be:

<http://ntvmr.uni-muenster.de/community/vmr/api/transcript/get/?docID=30093&bibliographyContent=Rev&fullPage=true&userName=darius.mueller&format=teiraw>

Notice the final noun/verb segments of the URL: transcript/get

²²⁶ In practice, only POST, GET, PUT, and DELETE are typically used in a REST implementation. For more on REST, see [section 3.4.6.2](#).

The individual functions (or “endpoints”) available in the VMR CRE are outlined and detailed below.

5.1.1. Authorization / Rights and Roles

Before any privileged services can be called (every service call which writes data into the system), a user of the API must first obtain a session token by logging into the system with the auth/session/open web service described below.

5.1.1.1. auth/session

The auth/session API group provides for login and session tracking within the VMR.

There are three actions available:

- open - opens a session, essentially logging into the VMR CRE
- close - closes a previously opened session
- check - checks the status of a session

The full URL to check a session at the NTVMR instance of the VMR CRE is:

<http://ntvmr.uni-muenster.de/community/vmr/api/auth/session/check/>

and sample output is:

```
<user id="13002" userName="tagriffitts"
sessionHash="5e6139b6-abc3-49df-96f1-40a34c9d3158">
  <emailAddress>scribe@crosswire.org</emailAddress>
  <firstName>Troy</firstName>
  <middleName>A.</middleName>
```

```
<lastName>Griffitts</lastName>
<jobTitle>Geek</jobTitle>
<userGroups count="3">
  <userGroup name="NT Transcripts Editors" userGroupID="272361"/>
  <userGroup name="Edwards Philemon" userGroupID="305870"/>
  <userGroup name="GSI-IGNTP 1 Timothy" userGroupID="598468"/>
</userGroups>
</user>
```

5.1.1.2. auth/hasrole

This authentication method can be used to determine if a user has the appropriate permission to execute a task in an installation of the VMR CRE. As an example to see if the current user has the role of “Index Manager”, the URL would be:

<http://ntvmr.uni-muenster.de/community/vmr/api/auth/hasrole/?role=Index+Manager>

and a result would be similar to:

```
<role name="Index Manager" userName="tagriffitts" hasRole="true"/>
```

5.1.2. Project and group management

5.1.2.1. projectmanagement/project

The project management functions provide access to project lists, information about those projects, and methods to manipulate those projects and members, given the proper access. Available methods include:

- list - list all projects available in the system
- get - get the details for a specific project
- put - create or modify a project
- delete - remove a project in its entirety from the VMR CRE instance

An example URL to list the projects within the NTVMR instance of the VMR CRE follows:

<http://ntvmr.uni-muenster.de/community/vmr/api/projectmanagement/project/list/>

5.1.2.2. projectmanagement/usergroup

The concept of a User Group in the VMR CRE revolves around a community with one or more projects. Typically a user group will have an independent community site within the VMR and any members of that community site are considered members of that user group. Methods available include:

- list - retrieve a list of all user groups available in an instance of the VMR CRE
- get - get the details for a specific user group
- ismember - checks if a user is a member of a specific user group

An example URL to list the user groups within the NTVMR instance:

<http://ntvmr.uni-muenster.de/community/vmr/api/projectmanagement/usergroup/list/>

To see if user *tgriffitts* is a member of the *Oxford Palimpsests* user group, the call would be:

<http://ntvmr.uni-muenster.de/community/vmr/api/projectmanagement/usergroup/ismember/?userGroupName=Oxford+Palimpsest&userName=tgriffitts>

5.1.2.3. projectmanagement/task

This API service provides for tasks to be assigned to project members. Methods available include:

- assign - assign a user to a task
- unassign - remove a user's assignment to a task
- get - retrieve the information about a task assignment, including progress completed
- list - list all tasks assigned to a user or project

An example URL showing the transcription progress of user *janneke* on manuscript 30001 for the ECM Matthew project:

<http://ntvmr.uni-muenster.de/community/vmr/api/projectmanagement/task/get/?projectId=1&objectID=30001&taskTypeID=3&assignedTo=janneke&detail=progress>

5.1.2.4. projectmanagement/project/data

The project/data facility provides a framework for projects to maintain a custom, versioned data repository. There are three methods available:

- put - adds or modifies data for a given key
- get - retrieves data for a given key
- listchildren - retrieves a list of all child keys under a given hierarchy

Project data keys can be defined hierarchically and a project can design any hierarchy it desires for its custom data repository, e.g., citations/latin/cyp/1. The listchildren method would give a list of all immediate children under any key, e.g., for citations/results might include: latin, greek, sahidic. Any data in any format may be stored with a key using the put/ endpoint and retrieved using the get/ endpoint.

5.1.3. Metadata

Metadata facilities provide services to describe and capture information about primary objects within the VMR CRE. These center around manuscripts and their folios, both as they existed in their original production unit, and also where they reside today.

Included is also an indexed set the portion of text which resides within these manuscript objects.

5.1.3.1. metadata/v11n (versification)

Versification describes the reference system used to refer to portions of the primary text of the edition and how we break that text up into research units. Within this software system, each largest unit must be defined as a named section, with 2 levels of

subdivision below this named section, each designated by a number in sequence with its siblings, e.g.,

Matthew	5	5
Antiquities of the Jews XII	2	16
Illiad	9	714
AntinoeB	1	6

Care must be taken when a project establishes a versification (v11n²²⁷); the v11n between published Bibles differs more than one might think. For example, the numbering of the Psalms in the Septuagint (LXX) is often one different to that of the Masoretic Text underlying most English translations. What are considered canonical Psalm titles are sometimes numbered alone, numbered as the first verse, or sometimes even as the first two verses of a Psalm, shifting the remaining verses in that Psalms out of alignment with other versification systems which do not include these titles in the reference system. Editions of the Septuagint, alone, display at least 20 different versification systems.²²⁸ In the New Testament, a similar example may be seen in occasional discrepancies between the Nestle-Aland Greek New Testament and the Stuttgart Vulgate, such as the last 20 verses of John 6.²²⁹ Many variations in

²²⁷ This abbreviation is used commonly at CrossWire to reduce typing. It is derived by removing the 11 characters between first and last character of the word “versification” in a similar vein to how i18n is commonly used as an abbreviation for “internationalization”.

²²⁸ CrossWire Bible Society, “LXX v11ns,” CrossWire SVN, 2018, http://crosswire.org/svn/sword-tools/trunk/versification/lxx_v11ns/.

²²⁹ For work done to harmonize different versification systems, see BiblIndex: Sources Chr tiennes, “Discordances in the verse numbering,” BiblIndex, 2017, <https://www.biblindex.info/en/biblical/list/discordance>; CrossWire under the subfolders here: CrossWire

versification can be handled by choosing a superset of closely matching editions which a project wishes to accommodate. A set of 15 supersets have been identified by CrossWire and used to digitize 1,447 texts spanning 797 languages. Much of this work has been done in partnership with Wycliffe Bible Translators, the American Bible Society, and the Society of Biblical Literature. The number of Bibles within each v11n system, determined by a scan of CrossWire’s digital library is listed below:

Calvin	1		LXX	5
Catholic	5		MT	9
Catholic2	1		NRSV	699
German	10		NRSVA	8
KJV	573		Synodal	56
KJVA	3		SynodalProt	52
Leningrad	7		Vulg	13
Luther	5			

Bible Society, “Versification,” Crosswire SVN, 2018, <http://crosswire.org/svn/sword-tools/trunk/versification/>; CCEL: Harry Plantinga, “CCEL/OSIS Reference Systems,” Christian Classics Ethereal Library (CCEL), 2017, <http://www.ccel.org/refsys/refsys.html>.

A custom versification for a text to be editing by the VMR CRE software system can be defined and registered with the system. The v11n definition should include accepted and preferred abbreviations and subdivision maximum numbers to assist the input parser with recognizing user input and ranges.

A v11n definition can be written as a text file following a simple XML structure exemplified as follows:

```
<v11n v11nID="LXX">
  <collection collectionID="1">
    <book osisID="Gen" name="1.Moses" preferredAbbrev="1.Mos"
      chapterMax="50">
      <chapter osisID="Gen.1" verseMax="31"/>
      <chapter osisID="Gen.2" verseMax="25"/>
      <!-- ... through chapter 50 -->
    </book>
    <book osisID="Exod" name="2.Moses" preferredAbbrev="2.Mos"
      chapterMax="40">
      <chapter osisID="Exod.1" verseMax="22"/>
      <chapter osisID="Exod.2" verseMax="25"/>
      <!-- ... through chapter 40 -->
    </book>
    <!-- ... all remaining books -->
  </collection>
  <!-- a second collections can be in included here, for organizational purposes
  only. used to separate New Testament books from Old Testament in the NTVMR -->
  <abbreviations>
    <abbr osisID="Gen">Genesis</abbr>
```

```

        <abbr osisID="Gen">Gen</abbr>
        <abbr osisID="Gen">1.Moses</abbr>
        <abbr osisID="Gen">1 Moses</abbr>
        <abbr osisID="Gen">I Moses</abbr>
        <abbr osisID="Exod">Exodus</abbr>
        <abbr osisID="Exod">Exod</abbr>
        <abbr osisID="Exod">2.Moses</abbr>
        <abbr osisID="Exod">2 Moses</abbr>
        <abbr osisID="Exod">II Moses</abbr>

        <!-- all other accepted abbreviations for each book referenced by osisID
        -->

    </abbreviations>
</v11n>

```

A versification system can be registered with the VMR CRE and set as the primary versification by including lines in the **sysconfig.properties** file such as:

AddVersifications=/home/vmr/v11ns/lxx.xml;/home/vmr/v11ns/masoretic.xml

PrimaryVersification=LXX

Multiple v11n systems can be registered with the AddVersifications directive by separating their definition file paths with a ‘;’ between, as seen in the example above.

Service calls for the metadata/v11n facility include:

- list - retrieve a list of registered versification schemes
- get - retrieve the definition of a versification scheme

- parse - ask the v11n parser to interpret user input against a given v11n system.

As an example, one can ask the parser to interpret the text “Gen3:5-4:9,12” against the LXXNU v11n system which comes as the default within the VMR CRE, returned as a list of contiguous verse ranges, with a call like:

<http://ntvmr.uni-muenster.de/community/vmr/api/metadata/v11n/parse/?v11nid=LXXNU&text=Gen.3:5-4:9,12&format=verses>

The verse parser does its best to identify partial words or partial abbreviations from the list of allowed abbreviation supplied in the v11n definition.

5.1.3.2. metadata/institute

The metadata/institute service facilitates the maintenance of a collection of holding institutes which house manuscripts or parts thereof which are included in the instance of the VMR CRE. Methods include:

- put - add or update an institution record
- get - retrieve an institution record
- getcountries - obtain a unique list of countries for which institution records exist in the system
- getplaces - obtain a unique list of places within a country for which institution records exist
- getnames - obtain a unique list of institution names within a country and place for which institution records exist in the system

For example, to see which institution records are available in London, UK, this call can be used:

<http://ntvmr.uni-muenster.de/community/vmr/api/metadata/institute/getnames/?country=United+Kingdom&place=London>

5.1.3.3. metadata/manuscript

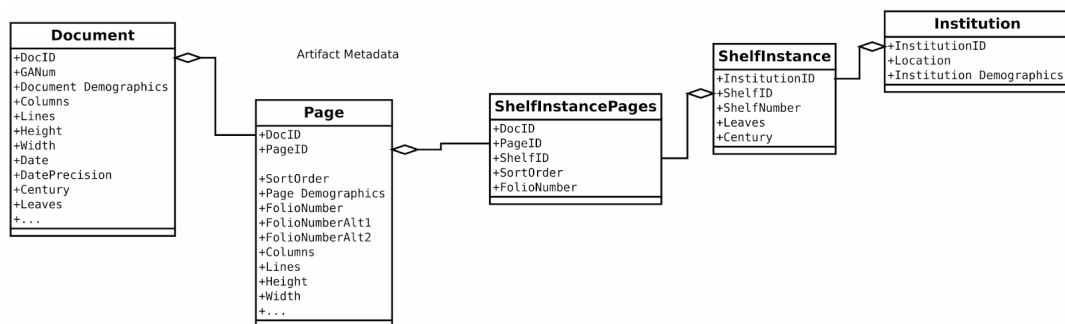
This service provides for the maintenance of manuscripts within the VMR CRE.

Available methods include:

- put - add or edit a manuscript record
- get - retrieve information about a manuscript
- putpage - add or edit a manuscript page
- deletepage - completely remove a manuscript page
- movepage - relocate data for a manuscript page
- clearnonhumanindexing - clear all indexing data associated with a manuscript which was added by automation (e.g., transcription import). This is useful when a mechanical import of a transcription was found to align incorrectly with manuscript pages in the VMR CRE and a re-import will be performed after adding more help markers (= folio numbers or verse numbers).
- clearfolionumbers - clear all folio numbers associated with a manuscript. See clearnonhumanindexing description for possible use cases.

5.1.3.4. metadata/shelfinstance

A shelf instance represents a library shelf mark which houses one or more manuscript pages. A shelfinstance object is associated with a holding institution. Manuscript pages are then associated with a shelfinstance object, established within the data model by the SHELFINSTANCEPAGES association table. This SHELFINSTANCEPAGES table bridges the gap between the concept of a “virtually reconstructed production unit” for a manuscript and the real-world contemporaneous situation with manuscript fragments residing at different libraries, museums, and private collections around the world, which is represented by Institution and ShelfInstance objects. The relevant excerpt from the data model is shown below.



Shelf Instance, Data Model Excerpt

5.1.3.5. metadata/liste (manuscript catalog)

The *liste* service is the main entry point for searches across the VMR CRE catalog of manuscripts. Methods include:

- get - get a brief summary of the catalog (or *liste*)

- search - a multifaceted search facility for the manuscript catalog. A number of parameters (or facets) can be combined to narrow results to precisely a desired set. Available parameters include:

biblicalContent	Results must contain some part of specified verses, [osisRef] e.g. "John", "John.1", "John.1.1"
docID	Limit results to single document [docID] e.g. 10046
gaNum	Limit results to documents matching Gregory-Aland number (= primary name), e.g. P46
pageID	optionally limit results to a single page [pageID], e.g., 10
lineCountMin	Limit results to documents with min line count [int]
lineCountMax	Limit results to documents with max line count [int]
columnsMin	Limit results to documents with min columns [int]
columnsMax	Limit results to documents with max columns [int]
pageWidthMin	Limit results to documents with min page width in cm [int]
pageWidthMax	Limit results to documents with max page width in cm [int]
pageHeightMin	Limit results to documents with min page height in cm [int]

pageHeightMax	Limit results to documents with max page height in cm [int]
dateMin	Limit results to documents with earliest possible date in year
dateMax	Limit results to documents with latest possible date in year
lang	Limit results to documents with specific language [string], e.g. grc
contentDesc	Limit results to documents with specific content overview, e.g. eap
featureCode	Limit results to documents with page feature code (parameter can repeat to widen the result set)
instID	Limit results to documents at a particular institute
shelfID	Limit results to documents in a particular shelf instance
detail	Level of result detail [(result count only), document, page]
limit	The approx. maximum number of pages to return. Will not return a partial document; any document started will complete and may exceed this limit. [(200)]
hasImages	[false true]
hasTranscriptions	[false true]

userName	owner of the document
-----------------	-----------------------

Examples:

The count of all manuscripts and pages which contain John 3:16 before AD 701:

<http://ntvmr.uni-muenster.de/community/vmr/api/metadata/liste/search/?biblicalContent=Jn.3.16&dateMax=700>

The same call adjusting the detail level to include each manuscript:

<http://ntvmr.uni-muenster.de/community/vmr/api/metadata/liste/search/?biblicalContent=Jn.3.16&dateMax=700&detail=document>

The same call including each page:

<http://ntvmr.uni-muenster.de/community/vmr/api/metadata/liste/search/?biblicalContent=Jn.3.16&dateMax=700&detail=page>

5.1.4. Object features

The feature concept in the VMR CRE allows for the dynamic extension of what research data will be captured for an object. The VMR CRE defines only a limited set of common metadata fields for manuscript and manuscript page objects. The Object

Features facility provides for the custom expansion to any data a research project may wish to capture. Features can be attached to either an entire manuscript or limited to a specific page of a manuscript. For example, one might wish to note that a manuscript has an alternate catalog identifier in another catalog system, or that an external image repository is known for a manuscript, or more locally, that a page contains an illumination. Features can be defined to simply mark an object as ‘having’ a feature, as in the latter example, or can hold data about a feature tag, as in the former example, what the catalog number actually is in the other catalog system.

The features service is broken down into 2 concepts. Each type of feature must first be defined before it can be used, this is done with the feature/definition service. Once the feature has been defined, it can then be used to tag any number of manuscripts or pages by using the feature/put method described in the next section. Features are referenced by a featureCode which uniquely identifies a feature definition. For more on the concept of Feature Tagging in the VMR CRE, see section [3.6. Tagging features](#).

5.1.4.1. feature/definition

The feature/definition service is used to define a feature so that it can be then used in the VMR CRE. The methods for this include:

- put - adds a new feature definition or updates an existing definition
- get - retrieves the definition of all features in the VMR CRE
- delete - completely removes the definition of a feature
- getvalues - retrieves a list of unique values for a feature tagged in the system

For example, to retrieve a list of all defined features, one could use this URL:

<http://ntvmr.uni-muenster.de/community/vmr/api/feature/definition/get/>

To retrieve all the unique values for the Grapheme features tagged in the VMR CRE, one would use:

<http://ntvmr.uni-muenster.de/community/vmr/api/feature/definition/getvalues/?featureCode=Grapheme>

5.1.4.2. feature

Once a feature has been defined using the feature/definition service described above, instances of the feature to tag manuscripts or on manuscript pages may be created using the feature service. Methods of this service include:

- put - create or update an instance of a feature tag
- get - retrieve instances of a feature tag
- delete - remove an instance of a feature tag

For example, to retrieve all feature tags associated with Manuscript 04, one would use this URL:

<http://ntvmr.uni-muenster.de/community/vmr/api/feature/get/?docID=20004>

5.1.5. bibliography

The Zotero bibliography system²³⁰ has been integrated into the VMR CRE, allowing an institutional Zotero account to manage all bibliographic material for their manuscript data. Each Zotero entry should be tagged with one or more Zotero tags which consist of the VMR CRE Document ID which is referenced by the bibliographic entry. The Zotero credentials for the institutional account should be entered in the **sysconfig.properties** file as such:

```
ZoteroUser=<user id number>  
ZoteroAPIKey=<user api key>
```

The bibliography service in the VMR CRE allows retrieval of references for all known writings about a catalogued manuscript. There is one ‘get’ call available. As an example, to see all bibliographic information available for P46 in the NTVMR, the call would be:

<http://ntvmr.uni-muenster.de/community/vmr/api/bibliography/get/?docID=10046>

5.1.6. image

The image service provides for the management of associated images with manuscript pages. There are 3 methods for this service:

²³⁰ Roy Rosenzweig Center for History and New Media, “Zotero | Your personal research assistant,” Zotero, 2018, <https://www.zotero.org>.

- put - add or modify a surrogate image for a manuscript page
- delete - remove an associated image from a manuscript page
- generatethumb - generate thumbnails for images.

These service calls all require elevated permission. The parameters which can be sent can be seen by browsing directly to the endpoint, e.g.,

<http://ntvmr.uni-muenster.de/community/vmr/api/image/put/>

5.1.7. biblicalcontent

This service supports the task of indexing pages within the VMR CRE. Methods include:

- put - add or modify indexing data for a page
- get - retrieve the indexing data for a page
- delete - remove the indexing data for a page

A sample call to see what indexing data is contained on the first page of Codex

Sinaiticus:

<http://ntvmr.uni-muenster.de/community/vmr/api/biblicalcontent/get/?docID=20001&pageID=40>

5.1.8. transcript

The transcript service provides facilities for working with transcriptions in the VMR CRE. Calls include:

- put - add or edit a manuscript page transcription
- get - retrieve a manuscript transcription for a particular page or verse range, in any one of a number of different formats.
- show - a service to retrieve a transcription which includes user controls to toggle between different viewing formats of the transcription
- clean - accepts a TEI transcription and removes extraneous attributes and tags and normalizes whitespace.
- getbestguesspage - given information about a transcription page, which may include folio number, verse content present on page, etc., this method will make a guess as to which page of a manuscript this information is associated. This is useful for importing transcriptions from an external source which does not include knowledge of VMR CRE pageID references.
- import - import a transcription from an external source
- import/form - a service which provides a very basic form to select and upload a transcription file to be sent to the import service
- splitpages - this service will split a transcription into discrete, autonomous, wellformed XML documents per page.
- search - retrieve information about all transcription work done by a particular user of the VMR CRE

Some examples:

Display the transcription of John chapters 8, 11, and 18-19 from manuscript 02, including the full page context, returned as XHTML:

<http://ntvmr.uni-muenster.de/community/vmr/api/transcript/get/?docID=20002&bibleContent=John.8,11,18-19&fullPage=true&format=xhtml>

Show the history of transcription for page 40 of Sinaiticus:

<http://ntvmr.uni-muenster.de/community/vmr/api/transcript/get/?docID=20001&pageID=40&history=true>

Show the version of that page created by lakmann on 2017-02-09 14:46:59, returned as TEI:

<http://ntvmr.uni-muenster.de/community/vmr/api/transcript/get/?docID=20001&pageID=40&versionHash=d0e6187f536ff82216620dbc8895813d56ed8f85&format=teiraw>

5.1.9. regularization

The regularization service provides for managing regularization rules typically defined during the editing process. A regularization rule can have a scope which applies the rule globally, or simply within a single verse. Methods available are:

- put - add or modify a regularization rule

- get - a rule or all rules appropriate for a context
- delete - remove a regularization rule
- import - import a rule set from an external source

For example, to obtain all the regularization rules which pertain to the context of Acts

1.4:

<http://ntvmr.uni-muenster.de/community/vmr/api/regularization/get/?indexContent=Acts.1.4>

5.1.10. documentgroup

The documentgroup service facilitates the management of user defined sets of manuscripts (a.k.a. “document lists” or “document groups”). Document groups are used by projects to limit the scope of their work, or simply as favorites lists for users to more quickly find frequently accessed manuscripts, or as a manuscript list to be used for a collation. Methods include:

- put - add or edit a document group
- get - retrieve a document group
- delete - completely remove a document group
- linkdocument - associate a document with a document group
- unlinkdocument - disassociate a document from a document group

A sample URL to retrieve all the document groups for a specific user would be:

<http://ntvmr.uni-muenster.de/community/vmr/api/documentgroup/get/?userName=tagri>
[ffitts](#)

5.1.11. collate

The collate service exposes the rich features of CollateX²³¹ coupled with the VMR CRE transcription repository. The service is complex and allows a large number of parameters to be specified. A full list of parameters follows:

[w<1-n>]	witness to include in collation. e.g., w1=text
[l<1-n>]	custom label for a witness. e.g., l1=P52
[format]	one of: (atable), graphml, dot, graph, tei, apptext, apphtml
[algorithm]	one of: (dekker), medite, needleman-wunsch

[documentGroupID]	collate witnesses which are members of a document group (-1 : all extant witnesses at a specific verse)
[indexContent]	which verse to collate (only applies to witnesses which are retrieved (e.g., from a documentGroupID))

²³¹ For more on CollateX, see [section 4.6.2.2](#).

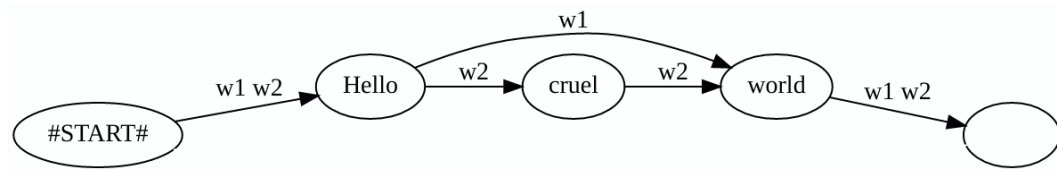
[baseText]	Base text module name to include in the collation, e.g., NA28
[loadModule]	(can be repeated) Additional module name to include in the collation, e.g., TR
[loadDocID]	(can be repeated) Witness to include in the collation, e.g., 10075
[ignoreSupplied]	true (false) Apply a general regularization to witnesses, ignoring supplied marks
[ignoreUnclear]	true (false) Apply a general regularization to witnesses, ignoring unclear marks
[ignorePunctuation]	true (false) Apply a general regularization to witnesses, ignoring Punctuation
[lang]	Restrict witnesses to a given language, e.g., grc

[regUserID[+*]]	apply regularization rules from this user, can be repeated, appended with '+' only localRules, appended with '*' only globalRules
[preferUser]	If set to a username and this username has a transcription

	for this verse, use the user's transcription over the global transcription
--	--

A simple example to retrieve a variant graph of the phrase, “Hello world,” collated with the phrase, “Hello cruel world”:

<http://ntvmr.uni-muenster.de/community/vmr/api/collate/?w1=Hello+world&w2=Hello+cruel+world&format=graph>



Variant Graph from collate/ Web Service

Other examples of calls to the collation service can be seen in the Referenceability section, specifically [sections 3.9.4](#) and [3.9.5](#).

5.1.12. Communication forum

5.1.12.1. forum/category

This service is used to manage topics or categories in the discussion forum. Methods include:

- assureexists - assure that a category exists in the forum. If it does not exist, create it.

- get - retrieve information about a category

5.1.12.2. forum/message

The forum/message service is used to retrieve from or create messages on the discussion forum within a specified category. Methods include:

- get - retrieve all messages within a specified category
- put - create a new message in a category

5.1.13. Test passages

5.1.13.1. variant/segment

Variant segments (or ‘test passages’) define the individual points in the text at which every manuscript will be examined by a project during the witness selection phase.

Calls to the variant/segment service include:

- put - add or update a variant segment for a project
- get - retrieve information about a variant segment or set of segments grouped by project

For example, to retrieve all the chosen test passages which the *Apokalypse* project in Wuppertal used for the witness selection phase:

<http://ntvmr.uni-muenster.de/community/vmr/api/variant/segment/get/?segmentGroupI>

[D=1](#)

5.1.13.2. variant/reading

A variant reading is an individual, unique reading at a given test passage. This service provides a facility to maintain the list of available readings from which a researcher must choose while examining a witness at a test passage. Calls include:

- put - add or update a variant reading option for a given variant segment
- delete - remove a variant reading from the list of options for a given variant segment
- reassign - this call will reassign all witness data from one reading to another.

This call is useful if a project decides that two readings are essentially the same and they wish to consolidate all choices made for two readings into one.

5.1.13.3. variant/apparatus

The variant apparatus endpoint can retrieve one segment of an apparatus at the specified verse location in the text. There is one call: get

For example to retrieve the edited apparatus data for Act 5:19:

<http://ntvmr.uni-muenster.de/community/vmr/api/variant/apparatus/get/?indexContent=acts.5.19>

The same, as HTML:

<http://ntvmr.uni-muenster.de/community/vmr/api/variant/apparatus/get/?indexContent=acts.5.19&format=html>

5.1.14. statistics

Statistic calls allow a client to retrieve information about the ongoing work and progress of work in an installation of the VMR CRE. Calls include:

- chaptercoverage - this call retrieves an overview of the witness coverage summarized by each chapter of a work.
- pages - a summary of all work done by manuscript page, categorized by type of work (cataloging, imaging, indexing, transcribing), and summarized by manuscript type.
- recentedits - this call retrieves a sorted list of most recent work done in the system, summarized by manuscript and includes the contributor with amount of recent work done.
- recentedits/rss - this call produces an RSS feed of similar data obtained by calling the recentedits service.

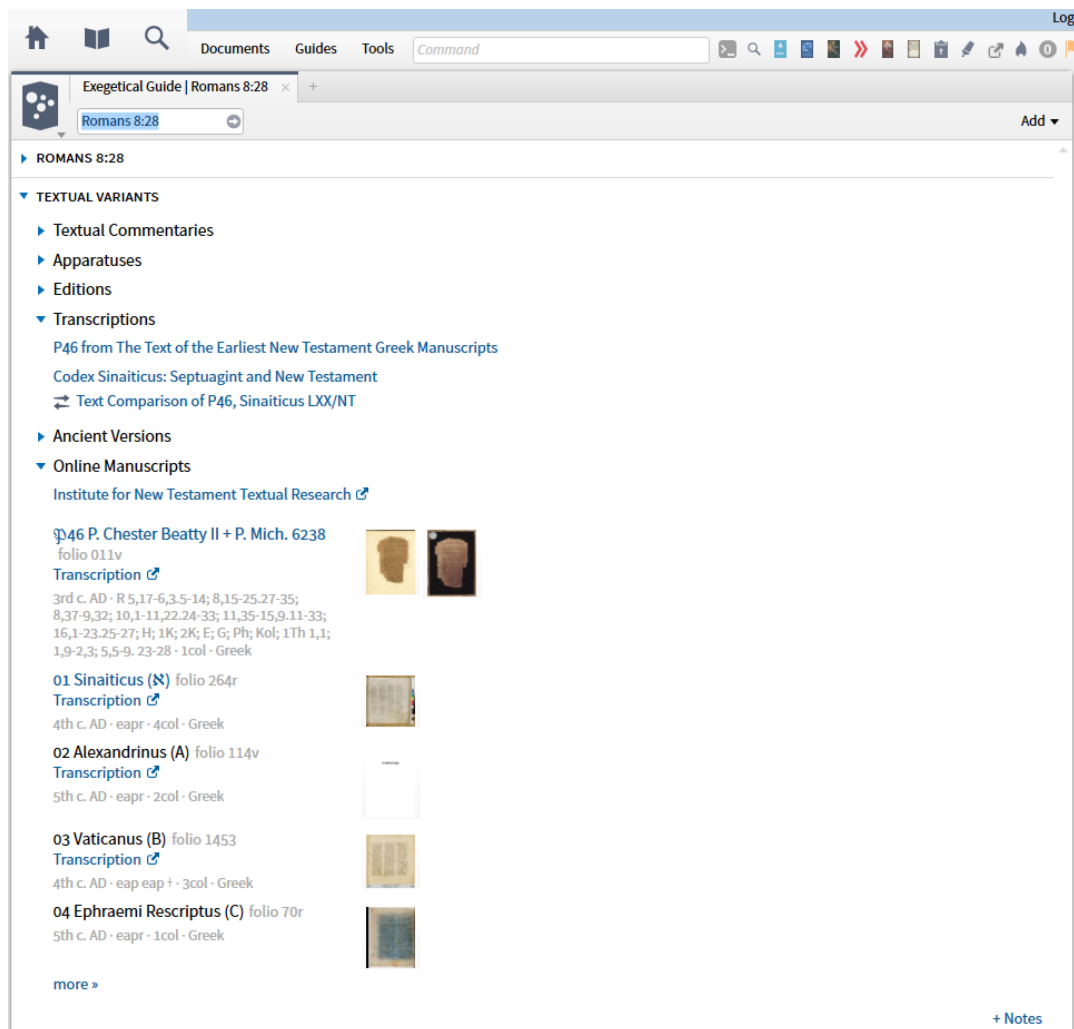
For example, to compute real-time data about all captured manuscript coverage by chapter of the New Testament,²³² one can use the call:

<http://ntvmr.uni-muenster.de/community/vmr/api/statistics/chaptercoverage/>

²³² A tool which provides a user display for this data is described in the following section on adoption.

5.2. Adoption

The VMR CRE web services API has achieved its purpose to expose the functionality of the software system and community-maintained dataset centered on New Testament Greek witnesses, as seen by the adoption and incorporation into other systems. Logos, the largest commercial Bible software publisher, has chosen to embed the functionality available through the web services API into their system.



Logos, Showing Integration with the VMR CRE via the API

CrossWire Bible Society, the largest non-commercial open source Bible software developer has also included various aspects of the software system into their offerings via the web services API. Their online research site, The Bible Tool,²³³ created for the American Bible Society and The Society of Biblical Literature, now includes a “Show Textual Evidence” button while performing a word study at any verse. This choice queries the manuscript evidence for the focused context from the INTF software system via the web services API and presents the results to the user, including direct links to transcriptions and images.

²³³ CrossWire Bible Society, “The Bible Tool: Parallel Viewing,” CrossWire, <http://crosswire.org/study/parallelstudy.jsp?del=all&add=Treg&add=NASB>.

Testing Your Faith

1 James, a bond-servant of God and of the Lord Jesus Christ,
To the twelve tribes who are dispersed abroad: Greetings.

2 Consider it all joy, my brethren, when you encounter various trials,
3 knowing that **the testing** of your faith produces endurance.
4 **And let endurance have** its perfect result, so that you may be
5 perfect and lacking in nothing.

wisdom, let him ask of God, who gives to all
reproach, and it will be given to him.

h without any doubting, for the one who
the sea, driven and tossed by the wind

to expect that he will receive anything from

d man, unstable in all his ways.

ble circumstances is to glory in his high

glory in his humiliation, because like
pass away.

a scorching wind and withers the grass; and the beauty of its appearance is destroyed; so amidst his pursuits will fade away.

perseveres under trial; for once he has received the crown of life which *the Lord* has reserved for Him.

he is tempted, "I am being tempted by God",
tempted by evil, and He Himself does not tempt

ted when he is carried away and enticed by

onceived, it gives birth to sin; and when sin is forth death.

my beloved brethren.

en and every perfect gift is from above,
father of lights, with whom there is no

dow.

will He brought us forth by the word of truth,
and of first fruits among His creatures.

1383 dokimion dok-im'-ee-on neuter of a presumed derivative of 1382; a testing; by implication, trustworthiness:--trial, trying. see GREEK for 1382

δοκίμιον

N-NSN

Search for lemma G1383 (any morphology) in Treg

Search for lemma G1383@N1-NSN (exact same morphology) in Treg

Search for lemma G1383 (any morphology) in LXX

Search for lemma G1383@N\-NSN (exact same morphology) in LXX

[View Image of Page 617 in Treg](#)

Some Manuscript Witnesses for Jas.1.3

Ms	Century	Folio	Content	
P74	VII	189	Jas 1:1-6	
01	IV	317v	Jas 1:1-27 Jas 2:1-6	
02	V	102r	Acts 28:30-31 Jas 1:1-15	
03	IV	1426	Jas 1:1-27	

This dataset is by no means exhaustive and is growing rapidly. Check back soon for more results.

Courtesy of [Institut für Neutestamentliche Textforschung](#)

The Bible Tool, Showing Integration with the VMR CRE via the API

A personal motivation for this present author to see the goal of this research project accomplished was to inform the conversations ongoing about the reliable transmission of the text of the Bible. Many inaccurate or misleading claims are made from all sides

of this debate. With the best-selling popular works from Dan Brown²³⁴ and Bart Ehrman,²³⁵ the lay conversation in and out of the church often includes statements ranging from, “How can you trust the Bible? It has been re-translated so many times through the years and modified by scribes, Roman emperors, and church councils to suit their needs,” to alternatively, “The New Testament has 24,000 ancient copies.”²³⁶ Both statements are, at best, misleading and the conversation needs to continue around a cogent and confirmable dataset. Until the dataset for this project was produced, it was impractical to obtain the extant manuscript coverage for every part of the New Testament, dated within a chosen range of history, and to confirm the evidence, if desired. Using the web services API, this data can easily be obtained and a tool²³⁷ to show the entire New Testament coverage by chapter, at a glance, for manuscripts between any selected date range has been developed to inform the conversation.

²³⁴ On page 195 in *The Da Vinci Code*, Brown write, “The Bible, as we know it today, was collated by the pagan Roman emperor Constantine the Great.” and “More than eighty gospels were considered for the New Testament.” Though fiction, this bestselling novel begins, “All descriptions of artwork, architecture, documents, and secret rituals in this novel are accurate.” (Dan Brown, *The Da Vinci Code*, 1. edition (New York: Corgi, 2004)).

²³⁵ Bart D. Ehrman, *Misquoting Jesus: The Story Behind Who Changed the Bible and Why*, Reprint (New York: HarperOne, 2007).

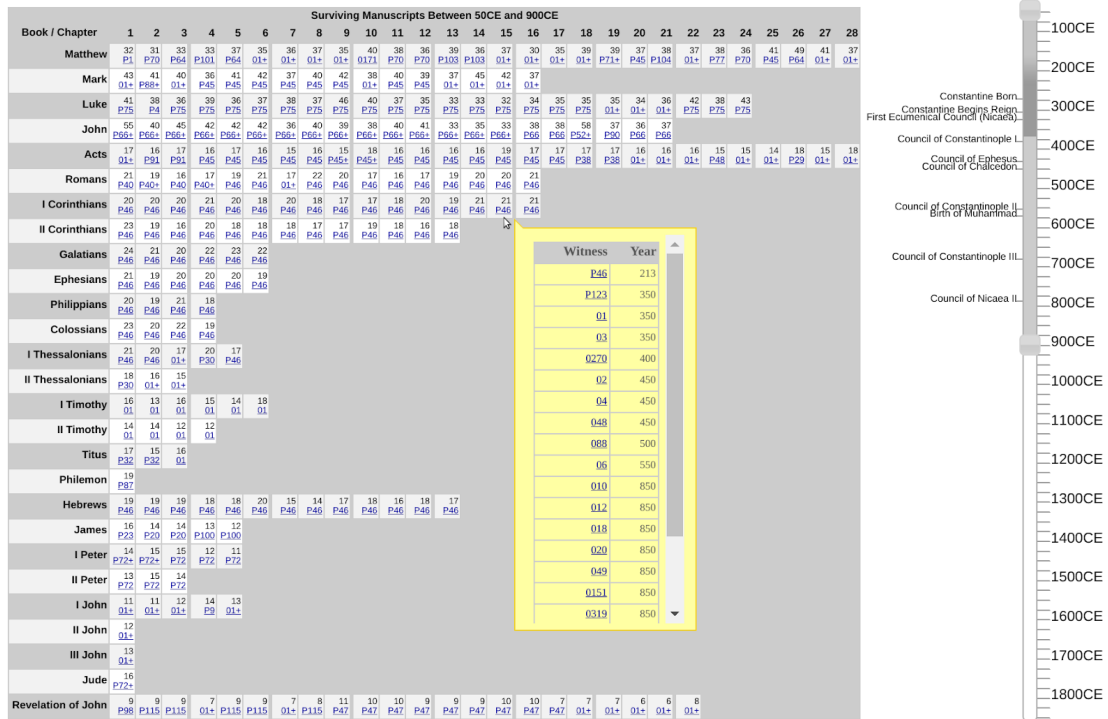
²³⁶ Encouraged by the popular works of Josh McDowell and others. (Josh McDowell, *The New Evidence That Demands a Verdict*, 1999 Edition: Fully Updated to Answer the Questions Challenging Christians Today, Rev ed.(Nashville, TN: Thomas Nelson Publishers, 1920)).

²³⁷ CrossWire Bible Society, “The Bible Tool: Manuscript Coverage,” CrossWire, 2018, <http://crosswire.org/study/rb/>.



THE BIBLE TOOL

[Home](#) [Passage Study](#) [Parallel](#) [Power Search](#) [Daily Devotionals](#) [Gospel Harmony](#) [Manuscript Coverage](#) [Preferences](#) [Help](#)



Greek New Testament manuscripts. Not exhaustive; showing manuscripts with INTF verse index data. Data courtesy of [The Institut für Neutestamentliche Textforschung](#)
Dates shown are an average of the earliest and latest estimates from the Institut für Neutestamentliche Textforschung.



Copyright © 2003-2017 [CrossWire Bible Society](#)

The Manuscript Coverage Tool, Using Realtime Data via the API

The timeline on the right includes important events in history. A user can drag the top and bottom handles of the timeline slider to change the date range, causing the table to update results accordingly. Each cell shows the number of witnesses extant today dated between the range selected, and also includes the oldest manuscript within that range.

If a user hovers over a cell, a full list of manuscripts will be presented which are extant for that chapter within the date range, shown along with their estimated copying date.

Any witness in the list can be selected to navigate the reader to that precise portion of the manuscript within the VMR CRE software system, showing both an image and a transcription, for reasonable verification. If desired, the user can even proceed from there to the “NT Transcripts” page in the software system to perform a realtime collation showing the exact variation between witnesses for any given verse. These resources greatly enhance the practical possibility to verify many of the statements, and to include verifiable reference data in new statements around this popular debate.

6. Beyond print

A primary objective for the printed critical edition, for centuries, has been to choose one form of a text to display and then to show selected variants in a manuscript tradition which are deemed important by the editor. Choosing the baseline reading is certainly a creative process. Additionally, choosing which units of variation to display against that chosen text form is a creative work. Deciding which sources are important enough to list as supporting evidence for each reading at a variant unit is also a creative work. Imagining how to present on a printed page the complex apparatus for an edition like the *Editio Critica Maior* is also a creative work. But are these choices what we, as readers of critical editions, value from the editor? Sometimes; but sometimes also, these are what a reader would like not to be in an edition. The ECM is not read lounging in the sun with an umbrella drink. It is a reference work. When a reader refers to the ECM, they wish to know, at a specific point in the text, which possible readings are present in the transmission tradition and which witnesses and other sources attest to a certain reading. This is the principal purpose for the ECM and with the goal to present an exhaustive apparatus instead of only selected points of variation, and with a body of evidence consisting of 150-200 chosen Greek manuscripts, biblical quotations from the first five centuries of Christianity, and evidence from multiple early biblical translations, this takes a substantial amount of time to produce. Where is the majority of that time spent? At the INTF, much of that time is filled, for example, with

deciphering a minuscule hand from a poor photograph of a manuscript page to produce a digital transcription. This task alone, done for the transmission tradition is an extremely valuable work to other researchers. Much of the editing time on those finished digital transcriptions is spent regularizing spelling, abbreviations, and other paleography into a regular form for comparison with other witnesses in their regular form. This regularization work is also a highly valuable resource to other researchers. The final product of a printed critical edition accomplishes what the data was intended *for*. In the digital world, we have the opportunity to present to the researcher what the data *is*-- all of it, as much or as little as the reader wants to see-- and leave it to the imaginations of the next generation what they might use the data for.

6.1. The digital edition of the *Editio Critica Maior*

The *Editio Critica Maior* is the paramount critical edition of the Greek New Testament. What features²³⁸ would a Digital ECM require to be deserving of this name? There are two mantra in software engineering that are important for our task at hand and one design principle that will guide our way. The software sage Joel Spolsky, in one of his popular articles writes on why it is never good to rewrite working and deployed code from scratch, “The idea that new code is better than old is patently absurd. Old code has been *used*. It has been *tested*. Lots of bugs have been found, and they’ve been *fixed*. There’s nothing wrong with it. It doesn’t acquire bugs just by sitting around on your

²³⁸ Patrick Sahle, Georg Vogeler, and the members of the IDE, “Criteria for Reviewing Scholarly Digital Editions, Version 1.1,” accessed August 28, 2017, <https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/>.

hard drive. Au contraire, baby! Is software supposed to be like an old Dodge Dart, that rusts just sitting in the garage? Is software like a teddy bear that's kind of gross if it's not made out of *all new material*?"²³⁹ This is wise commentary not only for software engineering, but also for our task to produce a digital edition of the ECM. The ECM, in its printed form, has been used for decades and follows a form that has been improved for centuries. It does not need redesign from scratch. Its readers have become accustomed to its nomenclature and methods, which have improved from user feedback over its lifetime. The second principle we will borrow from successful software engineering is authored by Max Kanat-Alexander as he prescribes the secret for maintaining a user base between software versions, "All you have to do to succeed in software is to consistently suck less with every release... Once [users have] picked a program, they will stick with it unless there is some compelling reason to leave. It's not like people constantly are looking for new software to replace yours-- they only start looking when your software just won't stop sucking."²⁴⁰ This is not to say that the printed edition of the ECM has flaws-- by all means, the editors have done wonders to present a treasure of information in the printed edition. But as the popularity of the digital medium grows, readers are expecting more. The point Kanat-Alexander makes is that a new revision must make some improvement on its previous edition to keep its current user base happy. Even with all the limitations of a printed edition, there is a real danger that a digital edition may suffer from being less convenient to use for a

²³⁹ Joel Spolsky, "Things You Should Never Do, Part I," *Joel on Software*, April 6, 2000, <https://www.joelonsoftware.com/2000/04/06/things-you-should-never-do-part-i/>.

²⁴⁰ Maxwell Kanat-Alexander, "The Secret of Success: Suck Less," *Code Simplicity*, August 11, 2009, <http://www.codesimplicity.com/post/suck-less/>.

readership-- a step backward for the reader. The Digital ECM must begin with at least the same convenience and also familiarity to which its print user base is accustomed. The Digital ECM should be an incremental improvement as it jumps into the digital realm. Finally, the French poet and aviator Antoine de Saint-Exupéry famously observed “Il semble que la perfection soit atteinte non quand il n’y a plus rien à ajouter, mais quand il n’y a plus rien à retrancher. Au terme de son évolution, la machine se dissimule.”²⁴¹ And so, in light of these three driving principles we shall develop the Digital ECM not redesigning the printed edition from scratch, but will take advantage of and preserve the improvements made to the edition over the many years of its existence; we will build the digital edition for a level of usefulness and convenience to at least that of its printed counterpart, to assure we are attractive and familiar to its current readership; and finally, we will focus not on adding to the design of the ECM, but instead on subtracting the limitations of the printed page.

Retaining the overall layout of the printed edition, we will begin with the familiar three panel design of the ECM page, keeping the eclectic edition text along the top with two columns of data beneath. The most simple improvement, moving from the printed page to a digital edition is the inclusion of scrolled data. This will enable us to restore logical context to the biblical text of the edition which was confined to an arbitrary window imposed by the physical limitations of the printed page. This is discussed in the following section.

²⁴¹Antoine de Saint-Exupéry, *Terre des Hommes* (Paris: N.R.F., 1939), <http://www.antoinedesaintexupery.com/terre-des-hommes-1939>.

6.1.1. Restoring context

The first limitation we will remove which is seen in the printed edition is the lack of context within the edition text. Implementing scrollable regions, we will restore two areas of lost context by first allowing the reader to browse an entire chapter of the text in a scrollable region along the top, and to also expand the apparatus material to cover any single verse in its entirety. The example included in the introduction to this research, showing only 17 words from James chapter II verse 3, now can include the apparatus for all of verse 3, and also make available to the reader for context, the entire chapter-- preceding and subsequent verses to our verse in focus. Scrollbars in the apparatus display also help with this task to remove the restrictions of the physical page, allowing the apparatus content for the entire verse to be included using only the left of our two columns beneath the edition text display. This leaves the right column available for our next improvement. Finally, to complete our familiar three panel layout, we will help restore linguistic context for the reader by incorporating a commonly used lexical aid, the Intermediate Liddell and Scott Greek-English Lexicon²⁴² from Perseus.²⁴³ A screen capture of the Digital ECM is shown below.

²⁴² Robert Scott, *An Intermediate Greek-English Lexicon: Founded Upon the Seventh Edition of Liddell and Scott's Greek-English Lexicon*, ed. H. G. Liddell (Oxford England: Benediction Classics, 2010).

²⁴³ Perseus Project, "Liddell and Scott's Intermediate Greek-English Lexicon (1889)," Perseus, 2018, <http://perseus.uchicago.edu/Reference/MiddleLiddell.html>.

NA28
Search
Quick Lookup: Acts 2:9

7 εξίσταντο δε και εθαυμαζον λεγοντες ουχ ιδου απαντες ουτοι εισιν οι λαλουντες γαλιλαιοι
8 και πως ημεις ακουομεν εκαστος τη ιδια διαλεκτω ημων εν η εγεννηθημεν
9 παρθοι και μηδοι και ελαμιται και οι κατοικουντες την μεσοποταμιαν ιουδαιαν τε και καππαδοκιαν ποντον και την ασιαν
10 φρυγιαν τε και παμφυλιαν αιγυπτον και τα μερη της λιβυης της κατα κυρηνην και οι επιδημουντες ρωμαιοι ιουδαιοι τε και προσηλυτοι
11 κρητες και αραβες ακουομεν λαλουντων αυτων ταις ημετεραις γλωσσαις τα μεγαλεια του θεου
12 εξίσταντο δε παντες και διηπορουν αλλος προς αλλον λεγοντες τι θελει τουτο ειναι

ECM
Realtime
Cit
VL
27 conjectures available.

2 a παρθοι ... 2147*. 2147Af

4 a και ... AnastSText. BasSel. Chrys. Cyr. Eus. GelCyz. LeontPrCP. PsCaes. SevGab. Socr. TertText. Thdrt

b om. 61. 326. 1838. 2495. 2718. L1825. AnastSMss. TertMss

— P74. P91. 025. 076. 095. 323S. 1739S. 1875. 1890S1. 2138. L156S

8-10 a και ελαμιται ... 01C2. AnastS. BasSel. Chrys. Cyr. Eus. GelCyz. LeontPrCP. PsCaes. SevGab. Socr. Tert. LzV. 5. 50. 51. 54. 56. 57. 58. 70. 72. 73. 189. K:SBM>. Å

b om. 01*

— P74. P91. 076. 095. 323S. 1739S. 1875. 1890S1. 2138. L156S

ΚΑΤΟΙΚΕΩ

I. to dwell in as a κάτοικος, to settle in, colonise, Hdt., Eur.: generally, to inhabit, Soph., Eur., etc.

2. absol. to settle, dwell, Soph., Eur., etc.:—so in perf. and plup. pass. to have been settled, to dwell, Hdt.

II. in Pass., of a state, to be administered, governed, Soph., Plat.

III. intr. of cities, to lie, be situate, Plat.

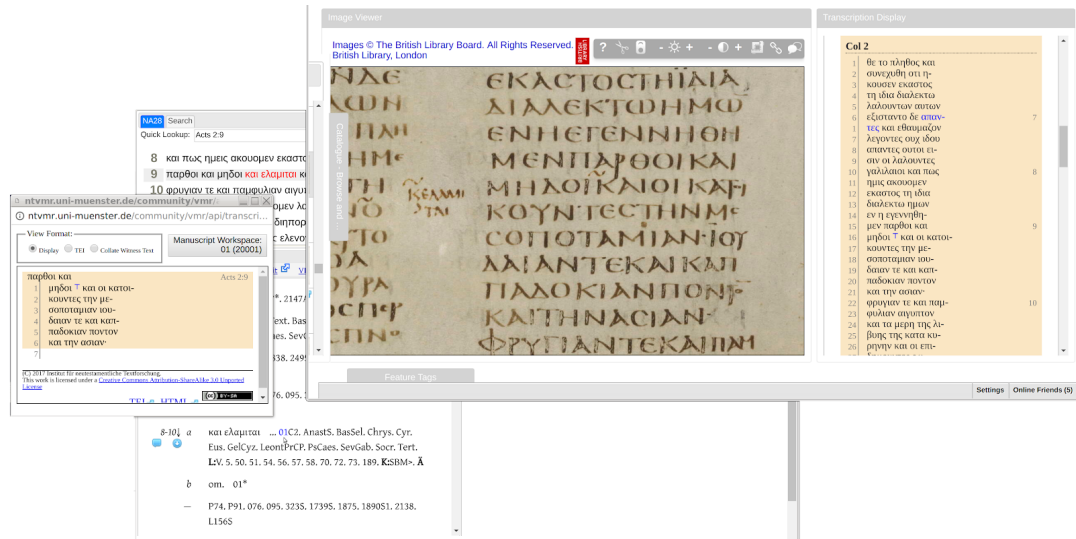
from Liddell and Scott, *An Intermediate Greek-English Lexicon*
Courtesy of [Perseus Digital Library](#)

Digital ECM: Acts

All of this information has been available in print; the digital edition merely makes access more convenient and immediately in view for the reader. We have only removed the time it takes to look up the material, and made space on the desk for a cup of coffee. The next section takes this one step beyond, also dealing with restoring context, but most of which was previously unavailable to the majority of ECM users before this present research project was implemented, and thus we will focus the next section more specifically on removing opacity.

6.1.2. A transparent conclusion

In previously printed editions of the *Editio Critica Maior*, the reader had to take the editors at their word concerning the evidence cited in the apparatus. Locating a transcription of the precise verse from an individual witness was often an impractical exercise. If one could find the source transcribed, the reader's trust then shifted from the apparatus editor to the transcription author. Locating and obtaining access to the physical object or appropriate images of the artifact to confirm that the transcription is accurate was unattainable to most. Now, with much of the ECM workflow facilitated within the VMR CRE software system developed as part of this research project, each transcription for every Greek witness cited in the ECM apparatus, along with indexed images of the artifacts themselves utilized in the production of those transcriptions are a slice of the readily available data which can be offered to the user of the Digital ECM. All Greek manuscript sigla listed in the apparatus now allow direct access to their respective verse transcription, which in turn is linked to an image of the precise folio side showing the verse in context directly on the artifact canvas.



Transcription and Image Display

When a Greek reading in the apparatus has been reconstructed from a versional source, the original source text can be shown when the mouse hovers over the reconstructed reading.

2298. 2344. 2412. 2495. 2652. 2718. 2805. 2818.
L156S. L1825. L2010. **Byz** [398ff2. 1739Sff1]
b τ (επι/εφ ημερας μ) 05*. 05C1f
c (επι/εφ ημερας μ) S:HM>
zwa/ao a/ao P56. Eus. OrLat
↔ a/b/c K:SBM
↔ a/c S:P>

Versional Evidence for Reconstructed Greek

Not only the internal data created during the production of the edition, but also external resources used for the research of the edition can be made convenient. The Digital ECM makes the *Vetus Latina* collations for Acts from the Universität Mainz, along with

the INTF's New Testament Patristic Citations, and also the Vrije Universiteit Amsterdam Database of New Testament Conjectural Emendation available for the verse in focus. Notice below the links across the top-right of the apparatus: "Cit", "VL", and "1 conjecture available" which display the three popup windows shown in the image.

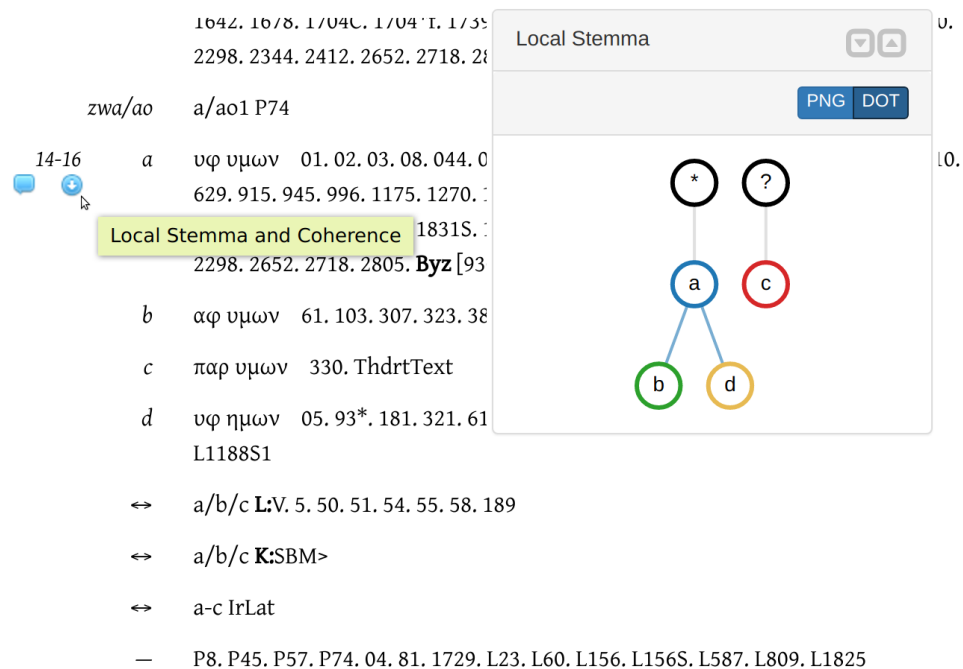
The screenshot displays the ECM interface for Acts 4:13. The main text area shows the Greek text of the verse. A sidebar on the left contains a table of variant readings. The bottom section features a table of patristic citations. A red circle highlights the links 'Cit', 'VL', and '1 conjecture available' in the top right of the main text area.

Link	Author	Writing	Chapter	Verse	Text/Context
8936	Call	VHyp	4	13	Εἰ ἤρην δι
5386	Chrys	Ant	4	13	τίς οὐκ ἂν αὐ
5667	Chrys	Bab	4	13	Οὐτε γὰρ λόγ
6674	Chrys	ExpPs	4	13	Πῶς γὰρ οὐκ
7007	Chrys	Hyp	4	13	ἄλλως δὲ

External Tools Connected

In addition to raw data resources, the editorial evaluation process-- the scientific tools used to make choices for the edition-- can also be reviewed by the reader. The process to evaluate thousands of variant passages via the Coherence-Based Genealogical Method for a fascicle of the ECM captures an enormous amount of data which reflects editorial decisions and this data is made immediately available to the reader. The Digital ECM provides, for any variant unit in the apparatus, an indication of the

availability of CBGM data, and if available, the ability to jump directly to that context within the data.



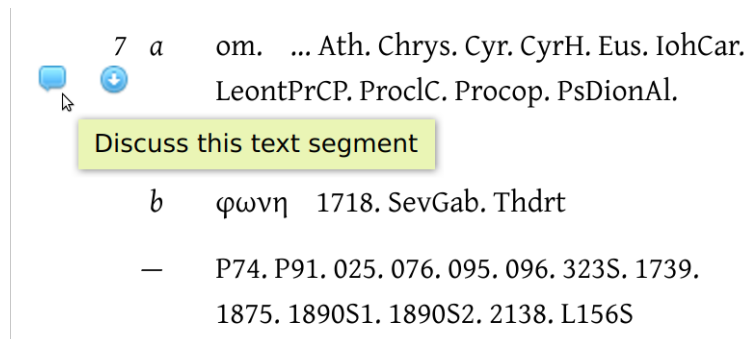
Local Stemma Shown from New Köln CBGM Framework

This new transparency brings the editors' *path to conclusion* immediately into view for peer review. No longer must a reader trust the summary of evidence in the apparatus or a transcription author, though they might and likely will most of the time, but the mere possibility alone to instantly check any conclusion in the edition raises the expected quality for the result-- twofold, first by instilling a new sense of fear in an originating editorial team with the new knowledge that, upon completion, every reader will be enabled to easily check their published work with the click of a mouse (of course this new sense of fear certainly does not apply to the editors of the ECM, all of whom always work with the highest integrity; naturally not those editors), and second, the

quality of the work perpetually improves with the combination of this continuous peer review by the reading community coupled with the ability to instantly publish a correction, inherent with a digital edition, which is the main topic for the next section.

6.1.3. Community: enhancement and perpetual revision

Scholarly dialog, in the world of print, happens in units of months. An edition will be published in August. Reviews will be commissioned in September and posted in October. A panel will meet to discuss the new edition in November. The digital edition has the opportunity to turn months into days, or even into realtime. The commission to build a thriving community of specialists around a digital edition is only to say that the common dialog which takes place now as a book review, a critical article, a response, should be encouraged and facilitated as part of the digital edition itself. The printed ECM already initiates scholarly dialog by including editorial commentary on selected passages. The Digital ECM uses this commentary as a springboard for discussion. First, facility for dialog has been added by offering a “Discuss” button as the reader’s mouse passes over any segment of the apparatus.



Discuss Button Highlighted

This tool will take the reader directly to a section within the ECM Textual Commentary online forum where discussion can be had for this context in the edition. The digital edition encourages the scholarly conversation by seeding the forum with the textual commentary from the printed edition for each context covered, as a thread of topic offered for discussion. The forum offers a subscription option to notify a reader when a new forum post is published-- any author of a post being subscribed to their own thread by default, to inform them of any replies. A snapshot of the edition and apparatus segment in focus is prepended to each forum message, along with a hyperlink to return a forum reader to the exact place in the edition on which a forum discussion is centered. This allows entry into the edition by way of either the discussion forum or the edition itself, and convenient navigation between the two.

[Message Boards Home](#) [Recent Posts](#) [My Posts](#) [My Subscriptions](#) [Statistics](#) [Banned Users](#)

← **Acts 1:10 (segment 34 - 36)**

Threads | [Previous](#) | [Next](#) | [Post New Thread](#) | [Permissions](#) | [RSS](#) | [Subscribe](#) | [Lock Thread](#) | [Move Thread](#) | [Move to the Recycle Bin](#)

Acts 1:10 (segment 34 - 36)

8/10/17 2:35 AM Status: **Draft**

[Reply](#) [Reply with Quote](#) [Quick Reply](#)

[Jump To ECM Entry](#)

Acts 1:10
και ως ατενίζοντες ησαν εις τον ουρανον πορευομενου αυτου και ιδου ανδρες δυο παρειστηκεισαν αυτοις εν εσθησεσιν λευκαις

εσθητι λευκη	P56*, 04C3.05.08.5.33.61.94.180.181.307.429.431.436.453.522.610.614.619.621.629.636.1162.1270.1292.
34-36 a	1297.1490.1501.1505.1595.1611.1678.1751.1838.1842.1875.1884.2138.2147.2200.2344.2412.2495.2652.2718.2805.2818.
	L60. L1825. L2010. Byz [424*, 886*, 1739Sf]. Chrys. OrLatMss. SevGab. ThdMop. S:PH
εσθησεσιν λευκαις	P56C*, 01.02.03.04*. 044.81.88*. 228.323.467.886K. 915*. 915Af. 945.996.1175.1409.1642.1704.1891.2298.
b	2774. L1178. Chrys(V). Cosmin. Epiph. Eus. OrLatText. L:54. 56. 58. 70. 72. 73. 189. K:SBM>
c εσθητι λαμπρα	1831S. Chrys
-- a/b	88Aff1. 424Cff2. 630ff1
-- a/c	L:5. 50. 51
-- b/c	Eus
--	P74. 319. 321. 323S. 1827S. 1890S2. 2570. L156S

GC: Coherence and diversity are good in the main strand of the attestation of *b* (starting with 03, if IR=?), but according to GC *b* emerged several times from *a*, while the attestation of *a* is perfectly coherent.

TP: The normalisation curve towards *b* (subject in plural) speaks for the priority of the majority reading *a*, the more so because the majority has the plural at Lk 24:4.

0 (0 Votes) [Flag](#) [Edit](#) [Top](#)

Discuss Post, Textual Commentary

This creates a home for the scholarly community to discuss the edition and more broadly the field of research. Removed is the distance in time between participant interaction. Access is granted for a wider audience of participation in the discussion. Over time the forum of discussion among scholars becomes an integral part of the digital edition to future readers-- with the potential to become as interesting as the original edition itself. As mentioned earlier, the discussion facility more simply also

provides a means for immediate feedback from a reader when an error is found in the edition, creating a perpetual cycle of review, feedback, and correction.

6.1.4. A dynamic research tool

The previous three sections have covered how a digital edition might bring added convenience for a reader, might more transparently expose the editor's path to conclusion by granting access to raw data and intermediate editorial decisions, and how the scholarly dialog around the edition might migrate to the digital realm. This section considers how the tools themselves which have been used for producing the edition might be made available for use by the audience. Many of the components built for the stages of editing a critical edition, discussed in [Chapter 4](#), lend themselves naturally to the reading audience. The manuscript catalog, with its multifaceted search, coupled with image and transcription viewing tools are straightforward choices. Feature tagging facilities, which enable the community to contribute metadata to the raw dataset behind the edition, provide a means of community enhancement to the project. But a primary component of a critical edition is the variant apparatus. How might this component transform from a static publication of the final conclusions of the editors, to a start to end, dynamic, realtime procession of the raw data through the rules and exceptions established by the editors to achieve the same apparatus display desired for the static print edition? Once this milestone is crossed, the digital edition will complete its transformation from a printed static publication of research conclusions, to instead the publication of an expert ruleset applied to a rich dataset-- a true expert system

encapsulating the expert knowledge of the domain masters in a field. The importance of this transformation from static to dynamic, from conclusions to meticulously defined methods as a ruleset, should not be understated. The result of the former is destined to be redone, as it has for centuries. But if the transition to a dynamic, realtime production of the edition display-- to the satisfaction of the editors-- is successful, then improving the data sources and tuning the rules gives way for the edition to perpetually advance into the foreseeable future. More data can be fed to the expert system, or possibly an entirely different dataset. The rules can be adjusted, new theories can be introduced into the ruleset and the results observed. The edition becomes a dynamic research tool breaking the repeated cycle of creating a new edition from scratch. The next generation of editors might truly stand on the shoulders of their predecessors.

Aside from the consideration of future editors, what advantages might a fully dynamic edition offer today's reader? First, there is further transparency offered with access to the editorial ruleset, along with the *assurance of adherence* to stated methods for the edition that comes with knowing the conclusions they view are generated from fully accessible raw data processed through fully accessible rulesets and algorithms. But what if the user of a digital edition could tune the ruleset, make different choices for which manuscripts should be included in the edition, have the option to visualize the results in different ways? The Digital ECM makes a first attempt at this goal by providing a "Realtime" tab on the apparatus display. Here, the reader can ask for a dynamic collation of the raw data through the rules established by the editors. A fully

dynamic edition has not yet been achieved. The dynamic result does not match that of the statically printed ECM apparatus. Only Greek witnesses are collated, excluding versional and other attestations. While global INTF regularization rules are applied to each transcription, the work to edit the Acts apparatus began before the Collation, Regularization, and Apparatus editing tools ([sections 4.6, 4.7](#)) were completed, thus no local regularization rules or variant settings are available to process for this feature. Still, the results are useful enough to make this a compelling tool for research and achieve the first steps toward a fully dynamic edition which produces the final results the edition editors wish to publish. Until this brass ring can be obtained, the Digital ECM for Acts presents both views: the static apparatus in its final edited form via the ECM tab, and also a Realtime tab which offers many tunable settings allowing the reader to experiment with realtime processing to achieve a dynamic result. This tool provides the opportunity for the researcher to begin asking questions of the data which the static apparatus might not answer. Seen below, the tunable parameters are multifaceted.

The screenshot displays the BibleWorks software interface. At the top, a search bar shows 'NA28' and a 'Quick Lookup' for 'Acts 1:12'. Below this, a list of verses in Greek is shown, with some lines highlighted in blue. The main window is divided into two tabs: 'ECM' and 'Realtime', with 'Realtime' currently selected. A menu is open over the 'Realtime' tab, listing options: 'Witness List', 'Base Text', 'Regularization Rules', 'Which Transcriptions', 'Include Extra Verses', 'Show Alignment Table', 'Show Variant Graph', and 'Download CSV'. The 'Realtime' tab displays a list of verse numbers and their corresponding Greek text. A 'Configure' button is visible on the right side of the interface. The bottom right corner shows a copyright notice: '(C) 2017 Institut This work is licen'.

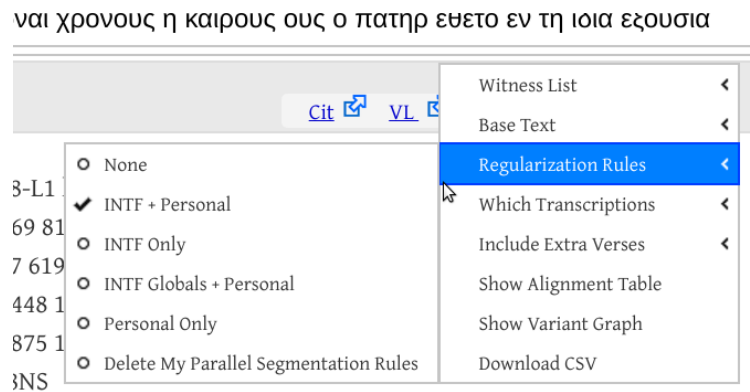
Realtime Tab and Menu Shown

First, the **Witness List** option allows the reader to tune the result by selecting which witnesses should be included in the processing. The ECM typically selects around 160 witnesses to include in its exhaustive apparatus, though the total witnesses cataloged for a book of the New Testament can be this number ten fold. Providing the option to select which witnesses to examine enables the reader to ask research questions like, “What variation exists *within* a single grouping of similar manuscripts? What variation exists *between* known groupings of manuscripts using one extant representative manuscript from each? How diverse were the manuscripts dated before the IV century?” User defined witness lists can be created in the Catalogue gadget (see

[section 4.1.2](#)) or the reader may select from the two virtual witness lists: INTF ECM Witnesses or All Extant Witnesses.

The **Base Text** option allows the reader to invert the critical apparatus, displaying the variation against any chosen witness in the witness list. This enables a reader to pivot from showing variation against the ECM reconstructed text, to displaying the textual tradition of variation against, for example, Codex Sinaiticus.

Regularization Rules provide for the researcher to select which regularizations should be performed against the manuscripts to be processed.



Regularization Rule Options

The reader may choose to turn regularization completely off and see every orthographical variation. The reader may include all the INTF regularization rules, augmented by their own personal rules, as seen selected by the check mark in the image above. A following paragraph will describe how the user is allowed to adjust the default word-level comparison within manuscripts by creating *ad hoc* spans of contiguous multi-word variation segments. The final option in the **Regularization**

Rules menu allows the user to clear these multi-word rules they have added for the current verse.

Transcriptions in the software system have three levels of production (described in [section 4.5](#)): personal (or initial), project (or reconciled), published. By default the Realtime processing tab in the apparatus display will process published transcriptions found within the system. The **Which Transcriptions** menu allows the user to choose other levels of publication, including the priority when multiple transcriptions exist for the same witness. Described earlier, for the Digital ECM, the scope of focus for the apparatus view has been re-contextualized to include one complete verse. Sometimes though, an interesting variant unit may span verses. The **Include Extra Verses** option expands the apparatus window to include the variation of up to four more verses.

Additionally, new experimental views for the results have been created. The familiar apparatus view is always displayed. The apparatus view suffers from the constraint that by definition, the results show each witness compared to a chosen baseline reading; the apparatus entries hang from a single text. Two added view options both remove this restriction by displaying the baseless collation (for a description of baseless collation, see [section 4.6.2.2](#)) which CollateX is able to perform. The first of these two views has been seen frequently in print²⁴⁴ and online publications as a means to show collation results: the alignment table. An alignment table records the entire verse from each witness transcription in a separate row of a table, then the words in each row are shifted to align parallel segments with each other.

²⁴⁴ Reuben Swanson has published a series of line-by-line collations, e.g., Reuben Swanson, ed., *New Testament Greek Manuscripts: Acts* (Sheffield: William Carey Int'l Univ Press, 1996).

be downloaded for offline use and opened in the reader's default spreadsheet application by selecting the **Download CSV** option.

The second alternative for display is the **Variant Graph**. A variant graph is similar to the alignment table in that it displays a baseless collation of all witnesses, read from left to right, but is different in that it collapses all identical readings within the same variant unit down into a single entry, showing only unique readings in the display. Readings are chained by connecting lines to show the reading path from start to end. Above each line are displayed the witnesses which follow that path through the text. The beginning of the results in the alignment table shown previously is displayed as a variant graph below.



Variant Graph

Notice all 20 manuscripts begin with *τοτε υπεστρεψαν*. The scribe named ‘corrector3’ in manuscript 04 has added the *οι αποστολοι* while the first hand of 04 and all other witnesses do not include this reading. The first hand of 044 has left out *εις ιερουσαλημ* but a corrector in that same manuscript has added it in. 08 is the lone witness to read *ιεροσολυμα* in place of *ιερουσαλημ*. All witnesses come back together and agree at *απο*. The variant graph display was originally developed as part of an Interedition focus group between Joris van Zundert, Tara Andrews, and this present author to design a multipurpose visualization and editing tool for both constructing regularization rules

and also establishing relationships between variants. Van Zundert developed the original implementation. Much of the original code has been incorporated into various tools including this one.

While the realtime processing capabilities in the software system have not yet achieved the full details of the final edited ECM apparatus, current capabilities already begin to realize the suggested benefits of that goal.

7. Summary of developments

In conclusion, compelling evidence for adopting the methods and practices proposed in this research can be seen concretely in the success of the VMR CRE software system and the community which has formed and is thriving around its NTVMR instance.

7.1. Contribution to the field

The INTF and partner institutes have adopted the system for much of their workflow. The ISBTF has published their *Text und Textwert* volume for the Apocalypse from data gathered using the Witnesses Selection module described in [section 4.2](#). The INTF is well on their way to completing the transcription and reconciliation editing phase for the book of Matthew for the ECM, managed by and conducted entirely within the software modules described in [section 4.5](#). An open data repository of the evidence for the Greek New Testament is now available online with many exciting new tools being developed which utilize the data services, with the most popular biblical research tools integrating these resources into their offerings (see, [API Adoption, section 5.2](#)). This open dataset now has a thriving community of scholars actively and naturally contributing to the dataset as they make use of the NTVMR for their own research. With over 1,500 registered users, the field of New Testament textual criticism has adopted the online platform as a central meeting point where open public conversations can be had regarding New Testament Greek manuscript data, showing 1,066 forum

messages in 752 threads, at the time of this writing. Houghton observes that “this is the pre-eminent site for the study of Greek New Testament manuscripts and continues to grow in both content and number of users. The initial adoption of shared electronic tools which facilitated the beginning of collaboration on editing the Greek New Testament has thus led to a new paradigm for the creation and publication of editions and their constituent data as well as the creation of a much larger community associated with these editorial endeavours.”²⁴⁵ The ECM volume for Acts is now successfully published online using the publishing module described in [section 6.1](#), constituting the first digital edition of the *Editio Critica Maior*. Beyond the primary focus for this research project which concentrates on the Greek New Testament, additional teams have added 188 Coptic and 11 Syriac New Testament manuscript resources and a project studying Climacus’ sixth century Ladder of Divine Ascent is conducting their work within the INTF’s system. New instances of the VMR CRE software system have been installed at the University of Göttingen to facilitate an online community for a 23 year *Akademie* project to produce a critical edition of the Sahidic Old Testament. The University of Berlin is researching the Avestan text tradition with their own instance of the software. The Austrian funded research project, The Canons of Apa John the Archimandrite, has chosen the software system to digitally publish their research, which has now been completed and made available online.²⁴⁶ Other academic

²⁴⁵ H. A. G. Houghton et al., “The *Editio Critica Maior* of the Greek New Testament: Twenty Years of Digital Collaboration,” *Digital Philology: A Journal of Medieval Cultures* 6, no. 2 (2017): Preview.

²⁴⁶ Diliaana Atanassova, “The Canons of Apa John the Archimandrite,” CoptOT, 2017, <http://coptot.manuscriptroom.com/web/apa-johannes/projekt>.

cataloguing efforts are integrating the services of the system into their work, including the ParaTexBib project.²⁴⁷

The various stages of collecting and processing data for an edition lends opportunity for scholars of varying skill levels to participate in the process. This participation is a pedagogical tool which helps the project, the student participant, and the professor teaching the subject. As students participate, they gain firsthand experience with the manuscripts, palaeography, the text, and the process of editing. A professor may choose to assign a manuscript indexing task of an appropriate level of difficulty to students of a beginning course on palaeography to familiarize students with an ancient hand. Experienced indexers may become beginning transcribers. An assignment to search for, and tag, paratextual features in a manuscript may help experienced readers learn ancient transmission forms of a text. The digital age brings opportunity for a new generation of students studying an ancient work to be exposed like never before to the elements of their research. This software platform encourages that exposure.

This case study serves as a blueprint from which other academic projects might pull ideas for their own work-- offered from a viewpoint which hopefully gives somewhat of a fresh perspective. Not only have software designs been suggested, but also lessons learned from building successful online communities, ways for a lone digital humanities software developer to build their development team in academia, and finally

²⁴⁷ ParaTexBib Project, "Paratexts of the Bible," ParaTexBib, 2014, <http://www.paratextbib.eu>.

yet another model for a digital scholarly edition (YAMDSE) has been offered.

As we have seen, New Testament research has been the focus of text-critical methods for centuries. Techniques ubiquitous in the humanities for establishing manuscript dates, transmission genealogies, and scribal hands and habits find their origins in New Testament text-critical research. With such a large body of material to process, theories and methods have historically been established using representative samples of the evidence. Modern technology now allows the comprehensive evaluation of the entire set of extant data-- if a normalized digital dataset is made available. This project has created the tools to allow this to happen, and at the end of this writing, 2.1M manuscript pages have been catalogued; 52% of these now have images available for viewing from anywhere in the world; only 10.21% or 219,172 pages have indexing information, and a meager 3.19% (67,905) have been transcribed. There is much work to do, but a goal has been set and the community has taken up the task. As this dataset grows, exciting new advances to legacy research-- statistically analysing palaeography to speak to questions of manuscript dating, text transmission families, and likelihood of scribal error-- by applying algorithms across comprehensive sets of data may tell us new information or give concrete data for old theories developed in times when it was not possible to perform comprehensive analysis.

7.2. Future considerations

We have discussed many of the benefits for doing research as a scholarly community. There are both new advantages and yet also problems that still need satisfactory solutions in this brave new world. Some further topics for future consideration are mentioned here.

It has been observed that adding a pedagogy element to a project can cause frustration for senior scholars “trying to get on with their work.” Consideration must be given to intentionally allot time for and prescribe as a project goal this pedagogical facet to alleviate this frustration-- avoiding the consideration only of the advantage (or disadvantage) a learning participant is bringing to the narrower goal to finish the edition. It is important to make clear from the outset that the task of training up the next generation of scholars is an essential and expected part of a senior researcher’s project contribution. This is a culture change for many in full time research.

Credit attribution in a collaborative community is not straightforward. Previously, in the ECM editing process, a transcription for an entire book was completed by a single transcriber. The term “a transcription” within the ECM editing world still often carries the definition “one text file of an entire (and only one) biblical book, transcribed (or reconciled) by one person.” In the world of community collaboration, volunteers come

and go. The task of transcribing an entire book is not attractive to most people.

Furthermore, in a flexible transcription repository, a query can be made for any range of data. A client might ask for the Synoptic Gospels, or simply John 3:16. Or the data for John 3:16 across 160 witnesses. An electronic transcription contributor may have corrected John 3:16. What credit should be shown? At what granularity should attribution be recorded? A manuscript feature is tagged by one user. A manager changes only one field of that feature tag and the original author disagrees. What attribution should be stored? How are conflicts resolved?

The digital edition is not yet fully dynamic, conclusions for Acts need to be reverse engineered to extract the regularization rules from the Greek apparatus. New concepts need to be developed moving away from associating local rules with word numbers, to something which caters better for an expert system which requires a level of fuzziness. Fuzziness scares the scholar who wants to be absolutely sure of exactly what happens, but without some level of fuzziness, artificial intelligence can never happen. Said differently, a ruleset cannot be applied to new datasets when a ruleset always specifies “only this manuscript, only this place.” A comfortable level of context needs to be included in a ruleset for patch theory to have a chance to succeed-- comfortable enough for the researcher to be happy that something reasonable will happen, yet fuzzy enough to allow the application of the rule to other data, e.g., “regularize this target word in John 3:16, as long as it is within the context of these 2 words before and this 1 word after the target word.” The context can be expanded or collapsed to adjust the comfort

level vs. the likelihood the rule can be applied to other data. Expectations need to be set for the scholar who is accustomed to publishing. A realtime option in a digital edition is not a publishing tool; it is a research tool. It gives results which the editor of the final edition might not want to see for various reasons. The choices the user selects might not make sense to the editor. The data might not be perfect yet in a community collaboration project. Rules might have been applied where they should not have been applied, due to an excessive selected level of fuzz. Much more work needs to be done to bring the realtime display to the point where it can achieve the results the editor desires for their choice of the final apparatus.

Two major components of the ECM editorial phase are not yet integrated into the system: the Versional and Citation editor developed at ITSEE and the new CBGM tools developed at the University of Köln.

To end on the note of collaboration and further, open development which has characterised this project as a whole, I will conclude by giving details of the locations from which the tools and open source code can be downloaded.

The Virtual Manuscript Room Collaborative Research Environment:

<http://vmrcr.org>

and the open source code:

<http://crosswire.org/svn/community/trunk>

My hope is that many others will find this a helpful environment in which to carry out their research and will supplement or fix anything according to their requirements and that a community might also grow from those who use and further develop these tools.

Bibliography

Adye, Willett L. *The History of the Printed Greek Text of the New Testament, with the Materials Available for Its Revision Considered: Being a Lecture Delivered at the Hartley Institution, Southampton, Jan. 30th, 1865, with a Supplement*. London: Rivingtons, 1865.

<https://archive.org/stream/historyofprinted00adyerich#page/n11/mode/2up>.

Aland, Barbara, Kurt Aland, Gerd Mink, Holger Strutwolf, and Klaus Wachtel, eds. *Novum Testamentum Graecum. Editio Critica Maior / Die Katholischen Briefe: Teil 1 und Teil 2, Gesamtwerk*. Stuttgart: Deutsche Bibelgesellschaft, 2013.

———, eds. *Novum Testamentum Graecum. Editio Critica Maior IV / Die Katholischen Briefe: Teil 1, 1. Lieferung, Der Jakobusbrief*. Stuttgart: Deutsche Bibelgesellschaft, 1997.

Aland, Kurt. *Computer-Konkordanz zum Novum Testamentum Graece von Nestle-Aland, 26. Aufl. und zum Greek New Testament, 3. ed.* Berlin: De Gruyter, 1963.

Aland, Kurt, and Barbara Aland. *The Text of the New Testament: An Introduction to the Critical Editions and to the Theory and Practice of Modern Textual Criticisms*. Leiden: Eerdmans and Brill jointly, 1987.

Aland, Kurt, Michael Welte, Beate Köster, and Klaus Junack. *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments*. Berlin: De Gruyter, 1963. <http://public.eblib.com/choice/publicfullrecord.aspx?p=3042078>.

Aland, Kurt, Klaus Witte, Barbara Aland, and Klaus Wachtel. *Text und Textwert der griechischen Handschriften des Neuen Testaments*. Berlin: De Gruyter, 1987.

Anderson, W.E. *Jane Austen's Novels as Represented Actions*. Berkeley, CA: University of California, 1968.

<https://books.google.de/books?id=XrxKAQAAMAAJ>.

- Andrews, Tara. "Digital Techniques for Critical Edition." In *Armenian Philology in the Modern Era*, 175–95. Leiden: Brill, 2014.
http://booksandjournals.brillonline.com/content/books/b9789004270961_008.
- Andrist, Patrick, Paul Canart, and Marilena Maniaci. *La Syntaxe Du Codex: Essai de Codicologie Structurale*. Bibliologia 34. Turnhout: Brepols, 2013.
- Barbrook, Adrian C., Christopher J. Howe, Norman Blake, and Peter Robinson. "The Phylogeny of The Canterbury Tales." *Nature* 394, no. 6696 (August 27, 1998): 839–839. <https://doi.org/10.1038/29667>.
- Barcan, Ruth. "Why Do Some Academics Feel like Frauds?" *Times Higher Education (THE)*, January 9, 2004.
<https://www.timeshighereducation.com/features/why-do-some-academics-feel-like-frauds/2010238.article#survey-answer>.
- Bons, Eberhard, and Jan Joosten. *Handbuch zur Septuaginta: Die Sprache der Septuaginta*. Gütersloh: Gütersloher Verlagshaus, 2016.
- Booch, Grady, Robert Maksimchuk, Michael Engle, Jim Conallen, Kelli Houston, and Bobbi Young Ph.D. *Object-Oriented Analysis and Design with Applications*. Boston, MA: Pearson Education, 2007.
- boss. "Feature Creep and the Importance of Iterations." *The MIT Entrepreneurship Review* (blog), December 1, 2010.
<http://mter.mit.edu/articlefeature-creep-and-importance-iterations/>.
- Brown, Dan. *The Da Vinci Code*. 1. edition. New York, NY: Corgi, 2004.
- Busa, Roberto, S.J. *Sancti Thomae Aquinatis Hymnorum Ritualium Vara Specimina Concordantiarum: A First Example of Word Index Automatically Compiled and Printed by IBM Punched Card Machines*. Vol. N. 7. Serie II. Milan: Fratelli Bocca, 1951.
- . "The annals of humanities computing: the Index Thomisticus." *Medicina Nei Secoli* 17, no. 3 (December 1980): 443–59.
- Chessell, Mandy, and Harald C. Smith. *Patterns of Information Management*. Upper Saddle River, NJ: IBM Press, 2013.
- Codd, E. F. "Normalized Data Base Structure: A Brief Tutorial." In *Proceedings of the 1971 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control*, 1–17. SIGFIDET '71. New York, NY: ACM, 1971.
<https://doi.org/10.1145/1734714.1734716>.

- De Hamel, Christopher. *The Book: A History of the Bible*. London: Phaidon Press, 2005.
- Dekker, Ronald Haentjens, Dirk van Hulle, Gregor Middell, Vincent Neyt, and Joris Van Zundert. "Computer-Supported Collation of Modern Manuscripts: CollateX and the Beckett Digital Manuscript Project." *Literary and Linguistic Computing* 30, no. 3 (September 1, 2015): 452–70.
<https://doi.org/10.1093/lc/fqu007>.
- DeRose, Steven. "Markup Overlap: A Review and a Horse." In *Proceedings of the Extreme Markup*. Montréal, Québec, 2004.
- Dinu, Valentin, and Prakash Nadkarni. "Guidelines for the Effective Use of Entity-Attribute-Value Modeling for Biomedical Databases." *International Journal of Medical Informatics* 76, no. 11–12 (2007): 769–79.
<https://doi.org/10.1016/j.ijmedinf.2006.09.023>.
- Dombrowski, Quinn. "What Ever Happened to Project Bamboo?" *Literary and Linguistic Computing* 29, no. 3 (September 1, 2014): 326–39.
<https://doi.org/10.1093/lc/fqu026>.
- Donovan, John. "Three-Tiered Client-Server Systems Gain as Enterprise-Wide Solution." *Investor's Business Daily*. October 10, 1994.
https://professordonovan.com/images/pdf/three_tiered_client_server_systems.pdf.
- Driscoll, Matthew James, and Elena Pierazzo. *Digital Scholarly Editing: Theories and Practices*. Cambridge, UK: Open Book Publishers, 2016.
- Ehrman, Bart D. *Misquoting Jesus: The Story Behind Who Changed the Bible and Why*. Reprint. New York, NY: HarperOne, 2007.
- Ehrman, Bart D., and Michael W. Holmes. *The Text of the New Testament in Contemporary Research: Essays on the Status Quaestionis. Second Edition*. Leiden: Brill, 2012.
- Ehrman, Bart D., Michael William Holmes, and Robert A. Kraft, eds. "The Use of Computers In New Testament Textual Criticism." In *The Text of the New Testament in Contemporary Research: Essays on the Status Quaestionis*, 268–82. Grand Rapids, MI: Eerdmans, 1995.
- Elliott, J. K. "Review of: Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments, IV Die Synoptischen Evangelien, 2 Das

Matthäusevangelium by Kurt Aland, Barbara Aland, Klaus Wachtel, Klaus Witte; Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments, IV Die Synoptischen Evangelien, 3 Das Lukasevangelium by Kurt Aland, Barbara Aland, Klaus Wachtel, Klaus Witte.” *Novum Testamentum* 41, no. 4 (1999): 390–92.

Finneran, Richard J. *The Literary Text in the Digital Age*. Ann Arbor, MI: University of Michigan Press, 1996. <http://trove.nla.gov.au/version/26561391>.

Fischer, Bonifatius. “The Use of Computers in New Testament Studies, with Special Reference to Textual Criticism.” *The Journal of Theological Studies* XXI, no. 2 (October 1, 1970): 297–308. <https://doi.org/10.1093/jts/XXI.2.297>.

Giachetti, Ronald E. *Design of Enterprise Systems: Theory, Architecture, and Methods*. Boca Raton: CRC Press, 2016.

Gurtner, Daniel, Juan Hernández Jr, and Paul Foster. *Studies on the Text of the New Testament and Early Christianity: Essays in Honour of Michael W. Holmes*. Leiden: Brill, 2015.

Harvey, Kate. “Feature Creep: What Causes It & How To Avoid It.” Chargify, July 12, 2016. <https://www.chargify.com/blog/feature-creep/>.

Houghton, H. A. G. “Codex Sinaiticus and Old Manuscripts of Early Christian Writing: Traditions and Innovations in Modern Research.” In *Sinajskij kodeks i pamjatniki drevnej christianskoj pis'mennosti: tradicii i innovacii v sovremennyh issledovanijach ; trudy Meždunarodnoj Naučnoj Konferencii “Sinajskij Kodeks ; Rukopis' v Sovremennom Informacionnom Prostranstve” ; (Pjatye Zagrebinskie Čtenija) ; Sankt-Peterburg, 12 - 13 nojabrja 2009 goda*, edited by Ekaterina Vladimirovna Krušel'nickaja, Rossijskaja Nacional'naja Biblioteka, Zagrebinskie Čtenija, and Meždunarodnaja Naučnaja Konferencija Sinajskij Kodeks: Rukopis' v Sovremennom Informacionnom Prostranstve, 38–49. Saint-Peterburg: Izdat. Rossijskoj Nacional'noj Biblioteki, 2012.

———. “Electronic Transcriptions of New Testament Manuscripts: The State of the Question.” presented at the Annual Meeting of SNTS (Studiorum Novi Testamenti Societas), South Africa, 2017.

———. “The Electronic Scriptorium: Markup for New Testament Manuscripts.” In *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, edited by Claire Clivaz, Andrew Gregory, and David Hamidović, 31–60.

Leiden: Brill, 2013.

<http://www.brill.com/products/book/digital-humanities-biblical-early-jewish-and-early-christian-studies>.

Houghton, H. A. G., D. C. Parker, Peter Robinson, and Klaus Wachtel. "The *Editio Critica Maior* of the Greek New Testament: Twenty Years of Digital Collaboration." *Digital Philology: A Journal of Medieval Cultures* 6, no. 2 (2017): Preview.

Houghton, H. A. G., and Catherine J. Smith. "Digital Editing and the Greek New Testament." In *Ancient Worlds in Digital Culture*, 110–27. Leiden: Brill, 2016. https://doi.org/10.1163/9789004325234_007.

"ISO/IEC 10646-1:1993 - Information Technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane." Accessed July 18, 2017. <https://www.iso.org/standard/18741.html>.

Jongkind, Dirk. *Scribal Habits of Codex Sinaiticus*. Piscataway, NJ: Gorgias Press, 2007.

Kanat-Alexander, Maxwell. "The Secret of Success: Suck Less." *Code Simplicity* (blog), August 11, 2009. <http://www.codesimplicity.com/post/suck-less/>.

Karrer, Martin, and Wolfgang Kraus. *Septuaginta Deutsch - Erläuterungen und Kommentare: Band 1 und 2*. 1st ed. Stuttgart: Deutsche Bibelgesellschaft, 2011.

Kevern, Rachel. "Transcribing Greek Minuscule Manuscripts." Monograph, March 23, 2011. <http://epapers.bham.ac.uk/774/>.

Kraus, Wolfgang, and Martin Karrer, eds. *Septuaginta Deutsch: Das griechische Alte Testament in deutscher Übersetzung*. 2nd ed. Stuttgart: Deutsche Bibelgesellschaft, 2009.

Kreuzer, Siegfried. *Handbuch zur Septuaginta: Einleitung in die Septuaginta*. Gütersloh: Gütersloher Verlagshaus, 2016.

Lembke, Markus, Darius Müller, Ulrich B. Schmid, and Martin Karrer, eds. *Die Apokalypse: Teststellenkollation Und Auswertungen*. Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments 6. Berlin: De Gruyter, 2017.

- Matha, Mahesh P. *Object-Oriented Analysis and Design Using UML: An Introduction to Unified Process and Design Patterns*. New Delhi: Prentice-Hall of India Pvt. Ltd., 2008.
- McDowell, Josh. *The New Evidence That Demands a Verdict, 1999 Edition: Fully Updated to Answer the Questions Challenging Christians Today*. Rev ed. Nashville, TN: Thomas Nelson Publishers, 1920.
- Mink, Gerd. "Contamination, Coherence, and Coincidence." In *The Textual History of the Greek New Testament: Changing Views in Contemporary Research*, edited by Klaus Wachtel and Michael W. Holmes, 141–205. Atlanta, GA: Society of Biblical Literature, 2011.
- . "Eine Umfassende Genealogie Der Neutestamentlichen Überlieferung." *New Testament Studies* 39, no. 4 (October 1993): 481–99.
<https://doi.org/10.1017/S0028688500011929>.
- . "Problems of a Highly Contaminated Tradition: The New Testament. Stemmata of Variants as Source of a Genealogy for Witnesses." In *Studies in Stemmata II*, edited by Pieter Th van Reenen, A. A. den Hollander, and Margot van Mulken, 5–85. Amsterdam: John Benjamins Publishing, 2004.
- Muellner, Leonard. "Annotations and the Ancient Greek Hero: Past, Present, and Future." *Comunicar* XXII, no. 44 (January 1, 2015): 46–53.
<https://doi.org/10.3916/C44-2015-05>.
- Myshrall, Amy C., Rachel Kevern, and H. A. G. Houghton. "IGNTP Guidelines for the Transcription of Manuscripts Using the Online Transcription Editor." Monograph, May 4, 2016. <http://epapers.bham.ac.uk/2161/>.
- Nielsen, Jakob. "How Long Do Users Stay on Web Pages?" Nielsen Norman Group, September 12, 2011.
<https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages/>.
- O'Hara, Robert J., and Peter M. W. Robinson. "Computer-Assisted Methods of Stemmatic Analysis." *Occasional Papers of the Canterbury Tales Project* 1 (1993): 53–74.
- Ott, Wilhelm. "Software Requirements for Computer-Aided Critical Editing." In *Editing, Publishing, and Computer Technology: Papers Given at the 20th Annual Conference on Editorial Problems*, edited by Sharon Butler and William P. Stoneman. New York, NY: Ams Press, 1988.

- Parker, David C. *An Introduction to the New Testament Manuscripts and Their Texts*. Cambridge, UK: Cambridge University Press, 2008.
- . *Manuscripts, Texts, Theology: Collected Papers 1977-2007*. Berlin: De Gruyter, 2009.
- . “Review: Kurt Aland, Barbara Aland, and Klaus Wachtel, Eds., in Collaboration with Klaus Witte, *Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments*, Vol. 4: Die Synoptischen Evangelien, No. 2: Das Matthäusevangelium, and No. 3: Das Lukasevangelium.” *TC: A Journal of Biblical Textual Criticism* 5 (2000).
<http://rosetta.reltech.org/TC/v05/Aland-et-al-ed2000rev.html>.
- . “The Text of the New Testament and Computers: The International Greek New Testament Project.” *Literary and Linguistic Computing* 15, no. 1 (April 1, 2000): 27–41. <https://doi.org/10.1093/lc/15.1.27>.
- . “Through a Screen Darkly: Digital Texts and the New Testament.” *Journal for the Study of the New Testament* 25, no. 4 (June 1, 2003): 395–411. <https://doi.org/10.1177/0142064X0302500401>.
- . “Was Matthew Written before 50 CE?: The Magdalen Papyrus of Matthew.” *The Expository Times* 107, no. 2 (November 1995): 40–43. <https://doi.org/10.1177/001452469510700203>.
- Parker, David C., and Klaus Wachtel. “The Joint IGNTP/INTF *Editio Critica Maior* of the Gospel of John: Its Goals and Their Significance for New Testament Scholarship.” Conference presented at the Annual Meeting of SNTS (Studiorum Novi Testamenti Societas), Halle, August 2, 2005. <http://epapers.bham.ac.uk/754/>.
- Parker, David C., Klaus Wachtel, Bruce Morrill, and Ulrich Schmid. “The Selection of Greek Manuscripts to Be Included in the International Greek New Testament Project’s Edition of John in the *Editio Critica Maior*.” In *Studies on the Text of the New Testament and Early Christianity*, edited by Daniel M. Gurtner, Juan Hernández Jr., and Paul Foster, 287–328. Leiden: Brill, 2015. https://doi.org/10.1163/9789004300026_015.
- Pierazzo, Elena. “A Rationale of Digital Documentary Editions.” *Literary and Linguistic Computing* 26, no. 4 (December 1, 2011): 463–77. <https://doi.org/10.1093/lc/fqr033>.

- Richard, Larry W. "A Closer Look: Text Und Textwert Der Griechischen Handschriften Des Neuen Testaments: Die Katholischen Briefe." *Andrews University Seminary Studies* 34, no. 1 (Spring 1995).
- Robinson, Peter M. W. "Current Directions in the Making of Digital Editions: Towards Interactive Editions." *Ecdotica* 4, n.d.
https://www.academia.edu/3313829/Current_Directions_in_the_Making_of_Digital_Editions_towards_interactive_editions.
- . "Electronic Editions for Everyone." In *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*. Cambridge, UK: Open Book Publishers, 2010.
<https://doi.org/10.11647/OBP.0008>.
- . "New Directions in Critical Editing." In *Electronic Text: Investigations in Method and Theory*, P.145-172. Oxford, UK: Oxford University Press, 1997.
<https://doi.org/10.1093/acprof:oso/9780198236634.001.0001>.
- . "Some Principles for Making Collaborative Scholarly Editions in Digital Form." *Digital Humanities Quarterly* 11, no. 2 (2017).
<http://www.digitalhumanities.org/dhq/vol/11/2/000293/000293.html>.
- . "The Collation and Textual Criticism of Icelandic Manuscripts (I): Collation." *Literary and Linguistic Computing* 4, no. 2 (April 1, 1989): 99–105. <https://doi.org/10.1093/lc/4.2.99>.
- . "The Ends of Editing." *Digital Humanities Quarterly* 3, no. 3 (2009).
<http://digitalhumanities.org/dhq/vol/3/3/000051/000051.html>.
- . "Towards a Scholarly Editing System for the Next Decades." In *Sanskrit Computational Linguistics*, 346–57. Lecture Notes in Computer Science. Berlin: Springer, 2009. https://doi.org/10.1007/978-3-642-00155-0_18.
- . "What Scholarly Editors Need to Help Us Make Sense Together in the Digital Age." *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1, no. 1 (2009).
<https://letterpress.uchicago.edu/index.php/jdhcs/article/view/64>.
- Sacco, Kathleen L. *Supporting Digital Humanities for Knowledge Acquisition in Modern Libraries*. Hershey, PA: IGI Global, 2015.
- Sahle, Patrick, Georg Vogeler, and the members of the IDE. "Criteria for Reviewing Scholarly Digital Editions, Version 1.1." *Institut Für*

- Dokumentologie Und Editorik* (blog). Accessed August 28, 2017.
<https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/>.
- Saint-Exupéry, Antoine de. *Terre des Hommes*. Paris: N.R.F., 1939.
<http://www.antoinedesaintexupery.com/terre-des-hommes-1939>.
- Sandberg, Russel. "The Sun Network File System: Design, Implementation and Experience." in Proceedings of the Summer 1986 USENIX Technical Conference and Exhibition, 1986.
- Schmid, Ulrich B. "Transmitting The New Testament Online." In *Text Comparison and Digital Creativity*, edited by Ernst Thoutenhoofd, Wido Th. Peursen, and Adriaan Weel, 189–206. Leiden: Brill, 2010.
<https://doi.org/10.1163/ej.9789004188655.i-328.73>.
- Schmid, Ulrich B., W. J. Elliott, and David C. Parker. *The Gospel According to St. John: Majuscles*. Leiden: Brill, 2007.
- Schmidt, Desmond. "Towards an Interoperable Digital Scholarly Edition." *Journal of the Text Encoding Initiative*, no. Issue 7 (November 12, 2014).
<https://doi.org/10.4000/jtei.979>.
- Schreibman, Susan, Ray Siemens, and John Unsworth. *A Companion to Digital Humanities*. Hoboken, NJ: John Wiley & Sons, 2008.
- Scott, Robert. *An Intermediate Greek-English Lexicon: Founded Upon the Seventh Edition of Liddell and Scott's Greek-English Lexicon*. Edited by H. G. Liddell. Oxford, UK: Benediction Classics, 2010.
- Shillingsburg, Peter L. *Scholarly Editing in the Computer Age: Theory and Practice*. Athens, GA: University of Georgia Press, 1986.
- Smith, Adam. *The Theory of Moral Sentiments*. London: A. Millar, 1761.
- Smithies, James. "A View from IT." *Digital Humanities Quarterly* 5, no. 3 (2011).
<http://www.digitalhumanities.org/dhq/vol/5/3/000107/000107.html#scheinfield2010>.
- Spencer, Matthew, Klaus Wachtel, and Christopher J. Howe. "Representing Multiple Pathways of Textual Flow in the Greek Manuscripts of the Letter of James Using Reduced Median Networks." *Computers and the Humanities* 38, no. 1 (February 1, 2004): 1–14.
<https://doi.org/10.1023/B:CHUM.0000009290.14571.59>.

- . “The Greek Vorlage of the Syra Harclensis.” *TC: A Journal of Biblical Textual Criticism* 7 (2002).
<http://rosetta.reltech.org/TC/v07/SWH2002/index.html>.
- Spolsky, Joel. “Things You Should Never Do, Part I.” Joel on Software, April 6, 2000.
<https://www.joelonsoftware.com/2000/04/06/things-you-should-never-do-part-i/>.
- Stephens, Ryan, and Ronald Plew. *Database Design*. Carmel, IN: Sams Publishing, 2000.
- Stiner, Scott. “Software Development Tips: Storyboard For Simplicity.” Forbes. Accessed April 9, 2018.
<https://www.forbes.com/sites/forbestechcouncil/2017/04/06/software-development-tips-storyboard-for-simplicity/>.
- Swanson, Reuben, ed. *New Testament Greek Manuscripts: Acts*. Sheffield: William Carey Int’l Univ Press, 1996.
- “TEI: Text Encoding Initiative.” Accessed August 25, 2017. <http://www.tei-c.org>.
- The Unicode Consortium. *The Unicode Standard, Version 10.0*. Mountain View, CA: The Unicode Consortium, 2017.
- . *Unicode Standard: Worldwide Character Encoding, Version 1.0*. Reading, MA: Addison-Wesley, 1992.
- Wachtel, Klaus, and Michael W. Holmes, eds. *The Textual History of the Greek New Testament: Changing Views in Contemporary Research*. Atlanta, GA: Society of Biblical Literature, 2011.
- Wasserman, Tommy, and Peter J. Gurry. *A New Approach to Textual Criticism: An Introduction to the Coherence-Based Genealogical Method*. Atlanta, GA: Society of Biblical Literature, 2017. <http://www.jstor.org/stable/j.ctt1whm8zr>.
- Williams, Sam. *Free as in Freedom: Richard Stallman’s Crusade for Free Software*. Sebastopol, CA: O’Reilly, 2002.
- Wisse, Frederik. *The Profile Method for the Classification and Evaluation of Manuscript Evidence, as Applied to the Continuous Greek Text of the Gospel of Luke*. Studies and Documents, v. 44. Grand Rapids, MI: Eerdmans, 1982.
- Wraabe, Wesley. “Collation in Scholarly Editing: An Introduction.” *Fill His Head First with a Thousand Questions* (blog), July 26, 2008.

<https://wraabe.wordpress.com/2008/07/26/collation-in-scholarly-editing-an-introduction-draft/>.

- Zundert, Joris van. "Author, Editor, Engineer — Code & the Rewriting of Authorship in Scholarly Editing." *Interdisciplinary Science Reviews* 40, no. 4 (October 2, 2015): 349–75. <https://doi.org/10.1080/03080188.2016.1165453>.
- . "The Case of the Bold Button: Social Shaping of Technology and the Digital Scholarly Edition." *Digital Scholarship in the Humanities* 31, no. 4 (December 1, 2016): 898–910. <https://doi.org/10.1093/lc/fqw012>.
- Zundert, Joris van, and Peter Boot. "The Scholarly Digital Edition 2.0." *Bibliothek Und Wissenschaft* 44 (2015): 141–152.
- Zundert, Joris van, and Ronald Haentjens Dekker. "Code, Scholarship, and Criticism: When Is Code Scholarship and When Is It Not?" *Digital Scholarship in the Humanities* 32, no. suppl_1 (April 1, 2017): i121–33. <https://doi.org/10.1093/lc/fqx006>.